

## REPORT DOCUMENTATION PAGE

AD-A265 026



1B. RESTRICTIVE MARKINGS

3. DISTRIBUTION/AVAILABILITY OF REPORT  
Approved for public release;  
distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)

AFOSR-TR-89-0509

6a. NAME OF PERFORMING ORGANIZATION

6b. OFFICE SYMBOL  
(if applicable)

7a. NAME OF MONITORING ORGANIZATION

Utah State University

AFOSR/NL

6c. ADDRESS (City, State and ZIP Code)

7b. ADDRESS (City, State and ZIP Code)

Utah Water Research Laboratory  
UMC 8200  
Logan UT 84322-8200110 Duncan Avenue, Suite B115  
Bolling AFB DC 20332-64488a. NAME OF FUNDING/SPONSORING  
ORGANIZATION8b. OFFICE SYMBOL  
(if applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

AFOSR

NL

AFOSR-89-0509

8c. ADDRESS (City, State and ZIP Code)

10. SOURCE OF FUNDING NOS.

110 Duncan Avenue, Suite B115  
Bolling AFB DC 20332-0001PROGRAM  
ELEMENT NO.PROJECT  
NO.TASK  
NO.WORK UNIT  
NO.

61102F

2312

A4

11. TITLE (Include Security Classification)

12. PERSONAL AUTHOR(S)

Dr. William J. Doucette

13a. TYPE OF REPORT

13b. TIME COVERED

14. DATE OF REPORT (Yr., Mo., Day)

15. PAGE COUNT

Final

FROM 8/16/89 TO 8/15/92

1/15/93

63

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

FIELD GROUP SUB GR.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A microcomputer based Property Estimation Program (PEP), utilizing MCI-Property, TSA-property and property-property correlations and UNIFAC derived activity coefficients, was developed to provide both experts and non-experts with a fast, economical method to estimate a compound's S, Kow, Pv, Koc, H, and BCF for use in environmental fate modeling.

93-12387



20. DISTRIBUTION/AVAILABILITY OF ABSTRACT

21. ABSTRACT SECURITY CLASSIFICATION

UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☒ OTIC USERS ☐

(U)

22a. NAME OF RESPONSIBLE INDIVIDUAL

22b. TELEPHONE NUMBER

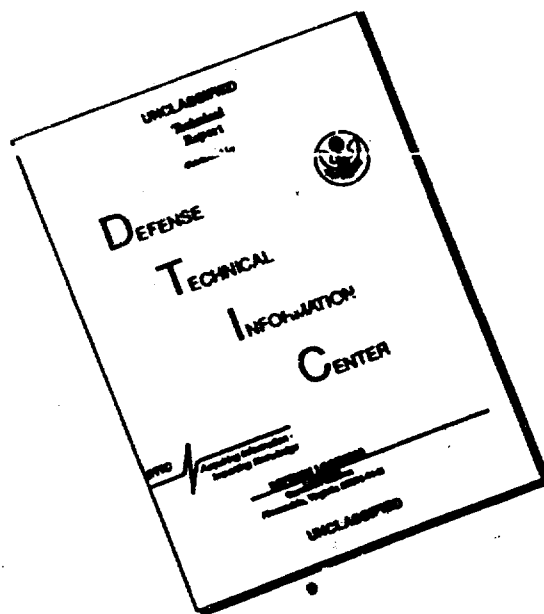
22c. OFFICE SYMBOL

Dr Walter J. Kozumbo

(202) 767-5021

AFOSR/NL

# DISCLAIMER NOTICE



THIS DOCUMENT IS BEST  
QUALITY AVAILABLE. THE COPY  
FURNISHED TO DTIC CONTAINED  
A SIGNIFICANT NUMBER OF  
PAGES WHICH DO NOT  
REPRODUCE LEGIBLY.

# ENVIRONMENTAL CONTAINMENT PROPERTY ESTIMATION USING QSARs IN AN EXPERT SYSTEM

William J. Doucette  
Mark Holt  
Doug Denne  
Utah State University  
Utah Water Research Laboratory  
Logan, UT 84322-8200

Final Report for Period 15 August 1991 - 15 August 1992  
Report No. AFOSR-89-0509

Prepared for  
U.S. AIR FORCE OF SCIENTIFIC RESEARCH  
Bolling AFB, DC 20332-6448

January 15, 1993

**DTIC QUALITY INSPECTED 4**

Accession For	
NTIS CRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

10 JAN 1993

# TABLE OF CONTENTS

	<u>Page</u>
Executive Summary . . . . .	1
Objectives or Statement of Work . . . . .	3
Background and Significance . . . . .	4
Status of Research Effort . . . . .	9
General Programming Description . . . . .	9
HyperCard . . . . .	9
Cards . . . . .	9
Stacks . . . . .	10
External Functions and Commands . . . . .	10
Algorithms . . . . .	11
TSA Algorithm . . . . .	11
MCI calculation method . . . . .	11
UNIFAC calculation method . . . . .	11
Fugacity level 1 model . . . . .	11
Development of QSPRs and QPPRs . . . . .	12
PEP Hardware/software Requirements . . . . .	12
PEP Overview . . . . .	12
PEP Processor . . . . .	13
MCI Module . . . . .	13
TSA Module . . . . .	18
Property/Property Module . . . . .	21
PEP Models . . . . .	23
PEP Help . . . . .	25
Chemical Property Database . . . . .	26
PEP Batch . . . . .	27
Summary . . . . .	32
Miscellaneous Publications . . . . .	33
List of Papers/Posters Presented at Professional Meetings . . . . .	33
List of Graduate Students Associated with the Research Effort . . . . .	34
References . . . . .	35
Appendix A: PEP user manual . . . . .	39
Appendix B: Statistical summary for MCI-property, TSA-property and property-property relationships incorporated into PEP . . . . .	53

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
1 The commercial XFCNs and XCMDs used . . . . .	10

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1 Flow chart illustrating the overall operation of PEP . . . . .	13
2 Flow chart describing operation of PEP MCI module . . . . .	14
3 Screen display of PEP MCI module . . . . .	16
4 Example statistics card from PEP . . . . .	16
5 Results card from PEP MCI module . . . . .	18
6 TSA module card from PEP . . . . .	19
7 PEP UNIFAC module . . . . .	21
8 PEP property/property correlation module . . . . .	22
9 Representation of Fugacity Level 1 Model compartments . . . . .	23
10. PEP Models card . . . . .	24
11 PEP Models results card . . . . .	25
12 Example Help card from PEP . . . . .	26
13 Example card from the Chemical Property Data Base . . . . .	28
14 PEP database degradation properties . . . . .	29
15 Example card for PEP Batch, MCI module . . . . .	30

## EXECUTIVE SUMMARY

Mathematical models are often used to estimate the fate and impact of organic chemicals in the environment. Use of these models requires a variety of parameters describing site and chemical characteristics. Aqueous solubility (S), the octanol/water partition coefficient ( $K_{ow}$ ), the organic carbon normalized soil/water sorption coefficient ( $K_{oc}$ ), vapor pressure ( $P_v$ ), Henry's Law constant (H), and bioconcentration factor (BCF) are considered key properties used to assess the mobility and distribution of a organic chemical in environmental systems.

One major limitation to the use of environmental fate models has been the lack of suitable values for many of these properties. The scarcity of data, due mainly to the difficulty and cost involved in experimental determination of such properties for an increasing number of synthetic chemicals, has resulted in an increased reliance on the use of estimated values.

Quantitative Property-Property Relationships (QPPRs), based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property. Quantitative Structure-Property Relationships (QSPRs) often take the form of a correlation between a structurally derived parameter(s), such as molecular connectivity indices (MCIs) or total molecular surface area (TSA) and the property of interest.

Selection and application of the most appropriate QPPRs or QSPRs for a given compound is based on several factors including: the availability of required input, the methodology for calculating the necessary topological information, the appropriateness of correlation to chemical of interest and an understanding of the mechanisms controlling the property being estimated.

Incorporation of QPPRs and QSPRs into a computer format is a logical and necessary step to gain full advantage of the methodologies for simplifying fate assessment. A Property Estimation Program (PEP), utilizing MCI-property, TSA-property and property-property correlations and UNIFAC-derived activity coefficients, has been developed for the Apple Macintosh microcomputer to provide the user with several approaches to estimate S,  $K_{ow}$ ,  $P_v$ , H,  $K_{oc}$  and BCF depending on the information available.

Structural information required for the MCI and UNIFAC calculation routines can be entered using either Simplified Molecular Identification and Line Entry System (SMILES) notation or connection tables generated with commercially available two-dimensional drawing programs. The TSA module accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by molecular modeling software. The program's built-in intelligence helps the user choose the most appropriate QSPR based on the structure of the chemical of interest. In addition, the statistical information associated with each QSPR in PEP can be displayed to help the user determine the model's validity. For the regression-based modules, assessments of accuracy based on the 95% confidence interval and estimated precision of the experimental values are also provided along with the estimated property value.

PEP also provides a batch mode that provides users with a method for the convenient, unattended calculation of MCIs, TSA and UNIFAC activity coefficients and the subsequent estimation of physical properties for large numbers of compounds.

A chemical property database, containing experimental values of S, Kow, H, Pv, Koc, and BCF compiled from a variety of literature sources and computerized databases was used for developing the MCI-property, TSA-property and property-property relationships used in PEP. This database, which currently contains over 800 chemicals, is linked directly to PEP.

The property estimation modules in PEP are also linked directly to the Level 1 and 2 Fugacity Models. The combination of the various property estimation methods, chemical property database, and simple environmental fate models provides users with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input.

## OBJECTIVES OR STATEMENT OF WORK

The primary goal of this project was to develop a microcomputer-based decision support system utilizing Quantitative Structure Property Relationships (QSPRs) and Quantitative Property Property Relationships (QPPRs) to predict the physical/chemical properties of an organic chemical which are necessary to model its environmental fate. The following specific properties were investigated: aqueous solubility (S), octanol/water partition coefficient ( $K_{ow}$ ), vapor pressure ( $P_v$ ), organic carbon normalized soil/water partition coefficient ( $K_{oc}$ ), Henry's Law constant (H), and bioconcentration factor (BCF).

In order to achieve the primary goal of this research, the following specific objectives were accomplished:

1. A database of experimentally determined values of S,  $K_{ow}$ ,  $P_v$ ,  $K_{oc}$ , H and BCF was compiled for over 800 organic compounds exhibiting a broad range of properties and expected mobility.
2. Algorithms to calculate molecular connectivity indices (MCIs), total molecular surface area (TSA) and UNIFAC activity coefficients were adapted/developed to run in a microcomputer environment using SMILES notation, connection files, or coordinate files to input required structural information.
3. Using the database described in Objective 1 and the computational methods developed in Objective 2, a variety of QSPRs and QPPRs for estimating S,  $K_{ow}$ ,  $P_v$ ,  $K_{oc}$ , H and BCF were developed.
4. Created a microcomputer-based decision support system that uses chemical structure information to aid the user in choosing the most appropriate QSPR or QPPR.
5. Linked property estimation routines and property database to simple environmental fate models ( the Level 1 and 2 Fugacity Models) to provides users with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input.



## BACKGROUND AND SIGNIFICANCE

Mathematical models are often used to estimate the fate and impact of organic chemicals in the environment. These models often idealize the environment as a system of connected compartments, i.e. water, soil, sediment, air and biota. The complexity of these models range from simple steady state models to non-steady state models which include a large number of compartments, transport between compartments and degradation processes.

Use of these models requires a variety of input parameters which describe site and contaminant physical-chemical and biological characteristics. Aqueous solubility (S), octanol/water partition coefficient ( $K_{ow}$ ), the organic carbon normalized soil/water sorption coefficient ( $K_{oc}$ ), vapor pressure ( $P_v$ ), Henry's Law constant (H), and bioconcentration factor (BCF) are considered key properties used to assess the mobility and distribution of a chemical in environmental systems.

One major limitation to the use of environmental fate models has been the lack of suitable values for many of these properties. The scarcity of data, due mainly to the difficulty and cost involved in experimental determination of such properties for an ever increasing number of synthetic chemicals, has resulted in an increased reliance on the use of estimated values. Quantitative Property-Property Relationships (QPPRs) and Quantitative Structure-Property Relationships (QSPRs) have been used by environmental scientists and engineers to obtain estimated values for a variety of physical/chemical properties for use in environmental fate and assessment modeling.

QPPRs, based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property without a specific concern for molecular structure. Frequently, the regression expressions are expressed in terms of the log of the two properties. Researchers have found that a number of environmental properties can be related to one another in this manner. For example, QSPRs have been developed to estimate S,  $K_{oc}$  and BCF from  $K_{ow}$  and  $K_{oc}$  and BCF from S [1-3].

QSPRs are methods by which the properties of a chemical can be inferred or calculated from a knowledge of the structure of a molecule. QSPRs often take the form of a correlation between a structurally derived parameter(s) and the property of interest. For example, relationships between structurally derived parameters, such as molecular connectivity indices (MCIs) and total molecular surface area (TSA) and properties such as S, Kow, BCF, and H have been reported.

Molecular connectivity developed by Randić [4] and refined and expanded by Kier and Hall [5-7] is a method of bond counting from which topological indices, based on the structure of the compound, can be derived. For a given molecular structure, several types and order of MCIs can be calculated. Information on the molecular size, branching, cyclization, unsaturation and heteroatom content of a molecule is encoded in these various indices [5]. MCI have been used to predict Koc [8,9], S [1], Kow [10], H [11] and BCFs [12].

A direct estimation of molecular surface area based on the concept of van der Waals radius, TSA has been correlated with S, Kow, Pv and H [13-22]. Several different algorithms, requiring the 3-D atomic coordinates of the solute molecule and the van der Waals radii of solute and solvent molecules as input [19,23], have been developed to calculate TSA.

Group contribution or fragment constant methods are another important category of QSPRs. The basic idea of a group contribution method is that while there is an enormous number of chemical compounds, both synthetic and naturally occurring, the number of functional groups that make up these compounds is much smaller. A single numerical value is assumed to represent the contribution of each functional group (i.e. a specified atom, a group of atoms bonded together or structural factor) to the physical property of interest. It is also usually assumed that the contributions made by each group are independent of each other. By summing up the values of the various fragments or groups the property of interest can be directly calculated.

The UNIFAC (UNIQUE Functional Group Activity Coefficient) group contribution method [24-26] has been used by environmental researchers to estimate S and Kow [27-31]. The UNIFAC method was developed to estimate liquid phase activity coefficients in mixtures of

nonelectrolytes [25]. In this technique, the activity coefficient is divided into two parts, a combinatorial part which reflects the size and shape of the molecule present and a residual portion which depends on functional group interactions. Various parameters, such as van der Waals group volumes and surface areas and group interaction parameters, are input into a series of equations from which the combinatorial and residual parts are calculated. Values for the group parameters have been tabulated and can be found in the literature[25,26]. UNIFAC is specifically designed to take into account interactions between groups and is appropriate for multiple solute/solvents systems. UNIFAC also permits estimates to be made as a function of temperature.

In most cases, more than one estimation method is available for a particular property. Estimation methods however, have widely varying accuracies and indiscriminate use of these techniques can result in large errors. Selection and application of QSPR or QPPR methods requires varying degrees of expertise that depend on the structure of a particular chemical of interest, knowledge of the mechanism of the process, the extent of the database used to develop the QSPR or QPPR and the complexity of the structural analysis required to relate structure to the property. For example, some QSPR and QPPRs are broader than others in the range of chemicals that are covered, and some methods have been established with a better understanding of the mechanisms or properties involved. In many cases estimation methods are developed from empirical or semiempirical correlations. The success of the correlation is dependent on many factors including the type and number of compounds used in its development.

Incorporation of QSPR and QPPRs into a computer format is a logical and necessary step to gain full advantage of the methodologies for simplifying fate assessment. A practical computerized property estimation program, utilizing QSPR and QPPRs, should include the following attributes: be simple and flexible to use for both experts and non-experts, include sufficient statistical information regarding the development of the QSPRs and QPPRs so that the range of applicability of such models can be evaluated, and provide an indication of the accuracy of the estimated property.

A microcomputer based Property Estimation Program (PEP), utilizing MCI-property, TSA-property and property-property correlations and UNIFAC derived activity coefficients, was developed to provide both experts and non-experts with a fast, economical method to estimate a compound's S, Kow, Pv, Koc, H, and BCF for use in environmental fate modeling. The user can input the required structural information for the MCI and UNIFAC calculation routines using either SMILES notation or coordinate files (connection table or "Molfile" formats) generated with commercially available two-dimensional drawing programs such as ChemDraw™ [45], Chemintosh™, or ISIS/Draw™. The TSA module accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by molecular modeling software such as Alchemy III™ or Chem3D Plus™. For property-property, TSA-property and MCI-property modules, the user can select from either "universal" or class specific regression models. To aid the user in choosing the most suitable regression model, the program automatically suggests the most appropriate regression model(s) based on the structure of the compound. In addition, the statistics associated with each model can be displayed along with the list of compounds used in developing the model. For the regression based modules, assessments of accuracy based on the 95% confidence interval and estimated precision of the experimental values are provided along with the estimated property value. Additional correlation models can be easily added to PEP by the user.

A chemical property database, containing experimental values of S, Kow, H, Pv, Koc, and BCF compiled from a variety of literature sources and computerized databases was used for developing the MCI-property, TSA-property and property-property relationships used in PEP. This database, containing over 800 chemicals, is linked directly to PEP and provides the means for the user to search for chemical compounds by full or partial name or synonym, to sort the compounds by name, boiling point, melting point,  $\sigma$ , molecular weight, and the ability to transfer to any of the property estimation modules.

In addition to the physical properties, the database was recently modified to allow the user to enter information pertaining to a compound's persistence and toxicity. Biodegradation rates, hydrolysis rates, photolysis rates and LC 50 values, along with the references and comments associated with each property, can be stored in the database.

To illustrate the potential application of PEP, the property estimation modules are linked directly to the Level 1 and 2 Fugacity Models developed by Mackay [32]. These simple models calculate the equilibrium distribution of an organic chemical between water, air, soil, sediment, suspended sediment and biota phases in a user defined world. The combination of PEP and Fugacity models provides users with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input. The development and use of the PEP system will be described.

## STATUS OF RESEARCH EFFORT

### General Programming Description

#### **HyperCard**

HyperCard is a program that was developed for the Apple Macintosh series of personal computers to enable novice Macintosh programmers to write user friendly computer applications. HyperCard, which is provided with every Macintosh sold, offers graphics, information storage, and the means to display information in a variety of formats. HyperTalk is a high-level, interpreted language used to establish links between related information and perform simple calculations within HyperCard. HyperCard also allows the programmer to create extensions of HyperTalk in a lower level language (i.e., C or Fortran). These extensions, called external functions (XFCN) and external commands (XCMD), greatly increase the speed of repetitive and calculation intensive algorithms over using HyperTalk itself. XFCNs and XCMDs can also be used to implement custom Macintosh features such as popup menus and custom dialog boxes.

#### **Cards**

Each screen of information in HyperCard is termed a card. Each card can contain graphics, fields, and buttons. The data on a card is held in the fields, and the buttons are used to initiate action procedures that operate on the data. The fields and buttons allow the standard Macintosh interface to be used without the direct use of the cumbersome Macintosh toolbox routines. To create a user interface the HyperCard programmer simply draws, or creates the buttons and fields. The link between buttons, fields and cards is done through HyperTalk scripts. A script is a set of HyperTalk statements linked to a button, field, card, or stack.

## Stacks

Cards are put together in HyperCard files called stacks. A stack can contain from 1 to 16,000 cards depending on the amount of memory each card requires. Usually each stack contains cards that are related either by purpose or visual similarity. The movement from stack to stack is rapid and easy to accomplish using either the standard Macintosh menus or HyperTalk scripts.

## External functions and commands

Some of the custom features used to enhance PEP were implemented using commercial XFCNs and XCMDs. Table 1 lists the commercial XFCNs and XCMDs that were used, their creator, and action.

Table 1. The commercial XFCNs and XCMDs used

XFCN or XCMD	Creator	Use
popUp	Adrian Freed (1989a) [42]	makes a pop up menu
ShowDialog	Jay Hodgdon (1988a) [40]	shows a modal Dialog
Progress	Jay Hodgdon (1988b) [41]	shows a dialog box with a progress pointer

The XFCNs and XCMDs that were used in PEP were created using Think C versions 4.0 and 5.0 from Symantec Corporation (1991) [44]. "Glue" routines are used to facilitate the communication between HyperCard and XFCNs or XCMDs. HyperCard glue is furnished with Think C. XTRA Shell by Adrian Freed (1989b) [43] was also used to develop XFCNs and XCMDs. XTRA also contains HyperCard glue plus a simple to use set of functions that can be called from Think C.

## Algorithms

### TSA algorithm

Total Surface Areas (TSA) are calculated using a modified version of the SALVO2 algorithm developed by Pearlman [16]. SALVO2, a FORTRAN program designed to run on main frame computers, was translated to the C computer language using Cobalt Blue's [46] FOR\_C translator version 2.9 (1989). This translation enabled the SALVO2 algorithm to be made into an XFCN and linked directly to a HyperCard stack.

### MCI calculation method

A C language program was written for the calculations of the MCIs based on code described by Frazier [35]. The algorithm currently calculates 54 (0 to 6 order) bond, valance, and path indices, and 7 (0 through 6 order)  $\Delta$  valence indices if the molecule contains any nitrogen or oxygen atoms. A more detailed discussion of the MCI calculation procedure is provided in the literature review section.

### UNIFAC calculation method

The UNIFAC procedure, as described by Grain [38], was incorporated into HyperTalk scripts and XFCNs. The group contribution factors were also taken from Grain [38] and are derived from vapor-liquid equilibria data.

### Fugacity level 1 model

The Fugacity level 1 model, described by Mackay [32, 39], was implemented in HyperTalk. This model is used to estimate the distribution of a chemical in a user defined environment consisting of a maximum of six compartments: air, water, soil, sediment, suspended solids, and biota. The default compartment volumes and densities were also taken from Mackay [32].



The graphs which show the distribution of the chemical are drawn using routines from GraphMaker a HyperCard stack included with version 2.0 of HyperCard. A detailed description of the PEP implementation of the Fugacity level 1 model is provided later.

### **Development of QSPRs and QPPRs**

The QSPRs and QPPRs utilized in PEP were developed using both statistical and intuitive criteria. The QSPRs were first derived using the stepwise regression features in StatView II, a statistical analysis package by Abacus Concepts Inc. (1988) [47]. The results from the stepwise regression procedure were analyzed and the variables containing theoretical information were left in the regression equation. The final regression equation was chosen to include both a size term and a measure of the polar nature.

After the regression equations were chosen the final calculations of the Analysis of Variance table and the graphs were obtained using Data Desk by Odesta Corporation (1989) [48]. Both universal and class specific equations for each property were developed and evaluated. All of the universal relationships and the class specific relationships that were found to be significant to the 90 percent level were incorporated in PEP.

### **PEP Hardware/software Requirements**

PEP requires the following system configuration to run: a Macintosh Classic, LC, II series, or PowerBook computer, with a hard disk; HyperCard 2.0 or greater software; Macintosh system software version 6.0.5 or greater, running under MultiFinder; and a minimum of 2 megabytes of memory (RAM), with 1000 kBytes of memory allocated for HyperCard.

### **PEP overview**

The PEP system currently consists of four HyperCard stacks: PEP Processor, PEP Models, PEP Help and Chemical Property Database. A flowchart illustrating the overall operation of PEP is provided in Figure 1.

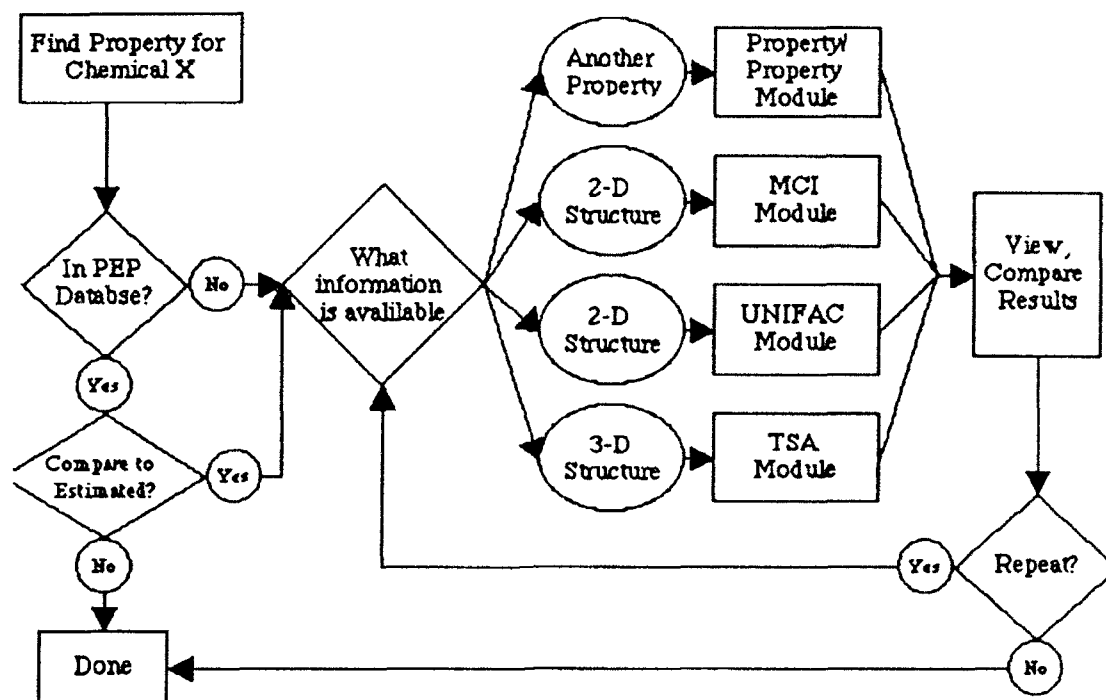


Figure 1. Flow chart illustrating the overall operation of PEP

Typically, users would first look for the required property information in the PEP database. If the information is not contained in the database, the user can then estimate the property using one or more of the four property estimation modules provided. Choosing the most appropriate property estimation module would depend on what information regarding the chemical is available. The function and use of each stack in PEP will be described in the following sections.

### PEP Processor

This stack, divided into four sections or modules, contains the algorithms for data input, calculations and output of the estimated physical-chemical properties. Each stack will be described in detail in the following sections.

**MCI Module** The overall operation of the MCI module is illustrated in Figure 2.

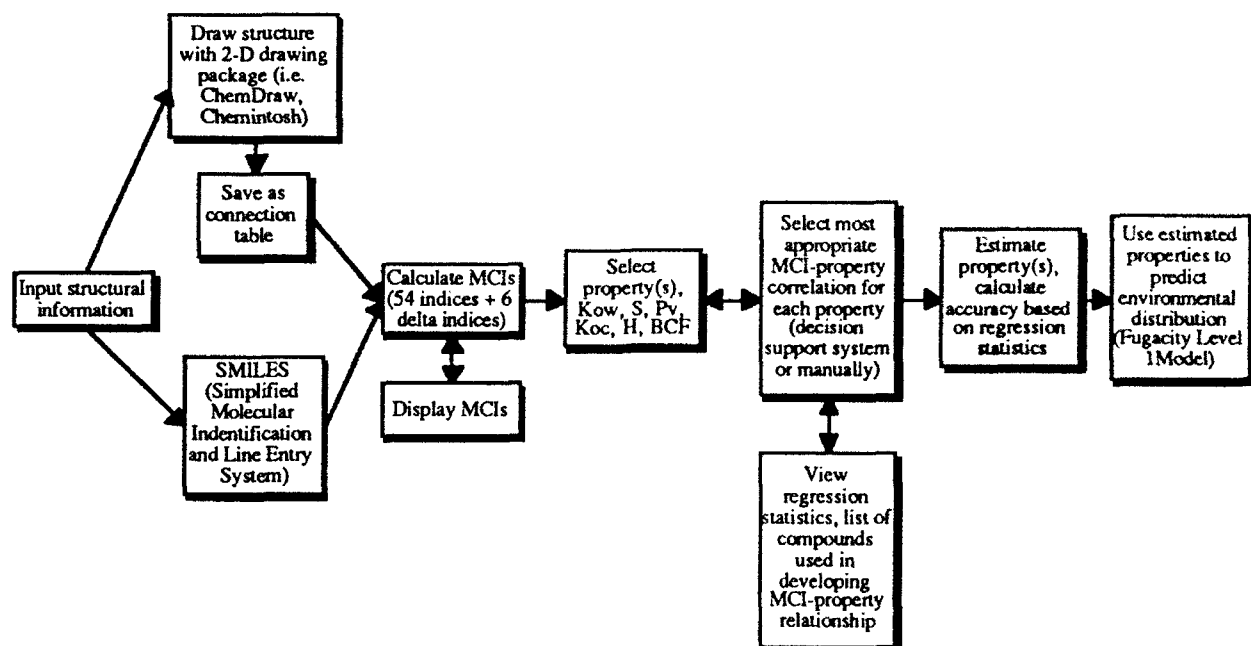


Figure 2. Flow chart describing operation of PEP MCI module.

The user interface of this module, shown in Figure 3, is designed in the form of a flow chart. Upon entering the MCI module the user must first input the necessary structural information using either SMILES [33,34] notation or connection files generated from ChemDraw™, Chemintosh™, or ISIS/Draw™, commercially available, Macintosh compatible two-dimensional (2D) drawing programs.

SMILES is a chemical notation language specifically designed for computer use. It is a method of "unfolding" a 2D chemical structure into a single line of characters containing the structural information.

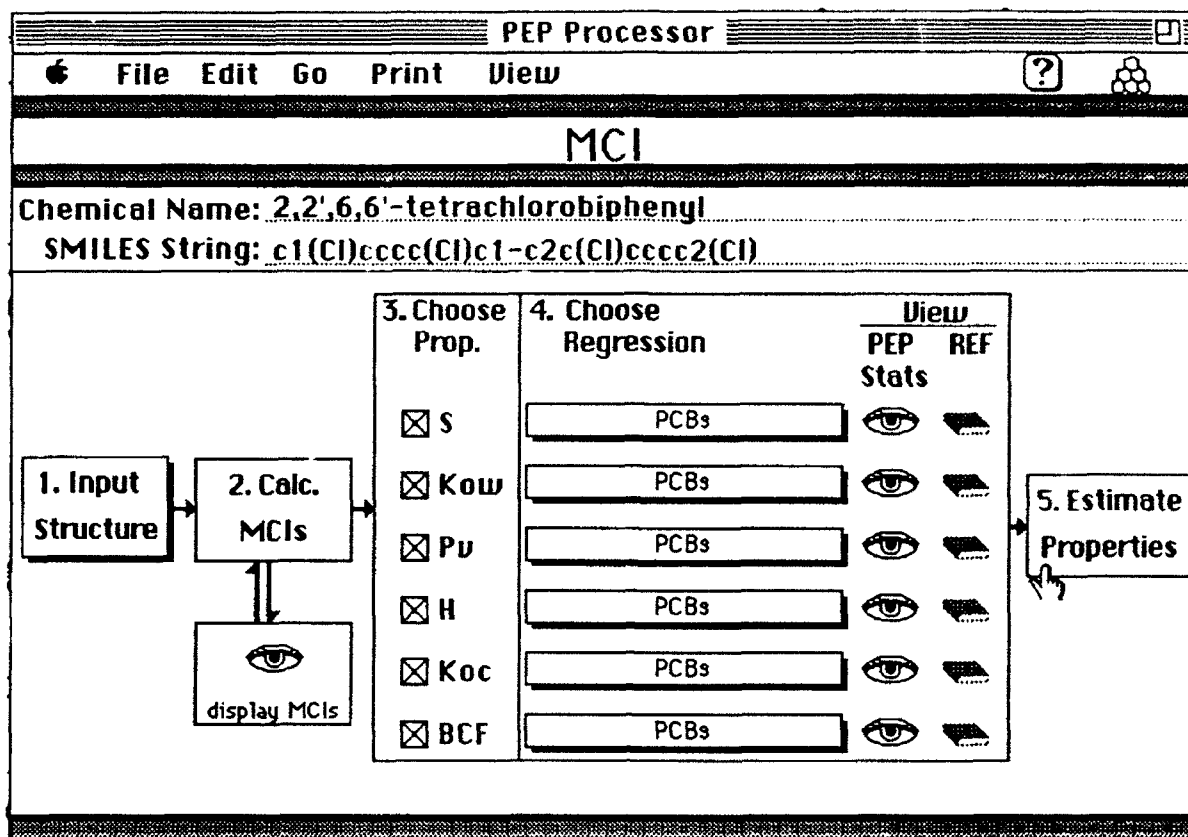


Figure 3. Screen display of PEP MCI module

After the structural information is entered, MCIs can then be calculated using a set of HyperCard™ external functions (XFCN) written in the programming language C based on code described by Frazier [35]. The MCI calculation routine in PEP calculates simple, bond and valence indices of several types (path, cluster, chain, and path/cluster) and orders (0 through 6), if possible, for each molecule, resulting in a maximum of 54 index values for each molecule which can be displayed on screen and/or output to a printer. To account for non-dispersive force effects on aqueous solubility and solubility related properties, zero through six order  $\Delta$  valence path indices ( $\Delta\chi$ ), as described by Bahnick and Doucette [36], are calculated by PEP, in addition to the 54 indices described above. To calculate  $\Delta\chi$  indices, a nonpolar equivalent is made by substituting C for O or N atoms. MCIs are calculated for the nonpolar equivalent and values for  $\Delta\chi$  can be computed for each type of index by:

$$\Delta\chi = (\chi)_{np} - \chi \quad (1)$$

After the MCIs are calculated, they can be displayed or printed if desired and the user can then choose which properties are to be estimated. For each property, two categories of MCI-property relationships are displayed. MCIs property relationships, both class specific and "universal", that were developed in this project using the experimental values reported in the PEP property database are preceded with the word PEP. "Universal" MCI-property relationships were developed using all available experimental data for a given property regardless of chemical class. "Class-specific" MCI-property relationships were developed if property values were available for a sufficient number (10 or greater) of compounds within a particular chemical class (PCBs, PAHs, ureas, etc.). In addition, several multi-class MCI-property correlations were developed for more broad classes of compounds such as: halogenated aliphatics and halogenated aromatics. An example illustrating the potential hierarchy of MCI-property relationships available to the user for the predicting the vapor pressure (Pv) of a polychlorinated biphenyl (PCB) is shown below. There are three MCI-property relationships, one developed using only PCBs, one using halogenated aromatics including PCBs and one using all compound types:

$$\log Pv = 5.814 (nc5) - 2.428 (np3) + 9.479 \quad (\text{PCBs})$$

$$\log Pv = -1.559 (bp1) + 6.622 \quad (\text{Halogenated aromatics})$$

$$\log Pv = -1.275 (np3) + 5.261 \quad (\text{Universal})$$

Generally, the use of a "class-specific" relationship, if available, should provide the best estimate (i.e. the estimate associated with the least amount of uncertainty).

By looking for a group of atoms and bonds that distinguish a chemical class, PEP uses the structural information contained in the SMILES string or connection file input to aid users in choosing the most appropriate MCI-property relationships. The number of appropriate relationships or chemical classes that are chosen by the program, denoted with a diamond in the popup menu, is determined by the number of different distinguishing subgroups that are found. In addition, a summary of the regression statistics and list of compounds used to develop and evaluate each MCI-property relationship can be displayed by clicking the "eye" or "view statistics option"

found at the left of each regression model. Information displayed on the statistics card includes: the MCI-property regression equation, the list of chemicals used in developing the regression model, the standard errors of the coefficients in the regression equation, the Analysis of Variance (ANOVA) table, the  $r^2$  value, a graph of the predicted vs. estimated values, a graph of the residuals vs. the predicted values, a graph of the residuals vs. the number of standard deviations and appropriate reference. An example of the statistical information provided for each MCI-property relationship is shown in Figure 4.

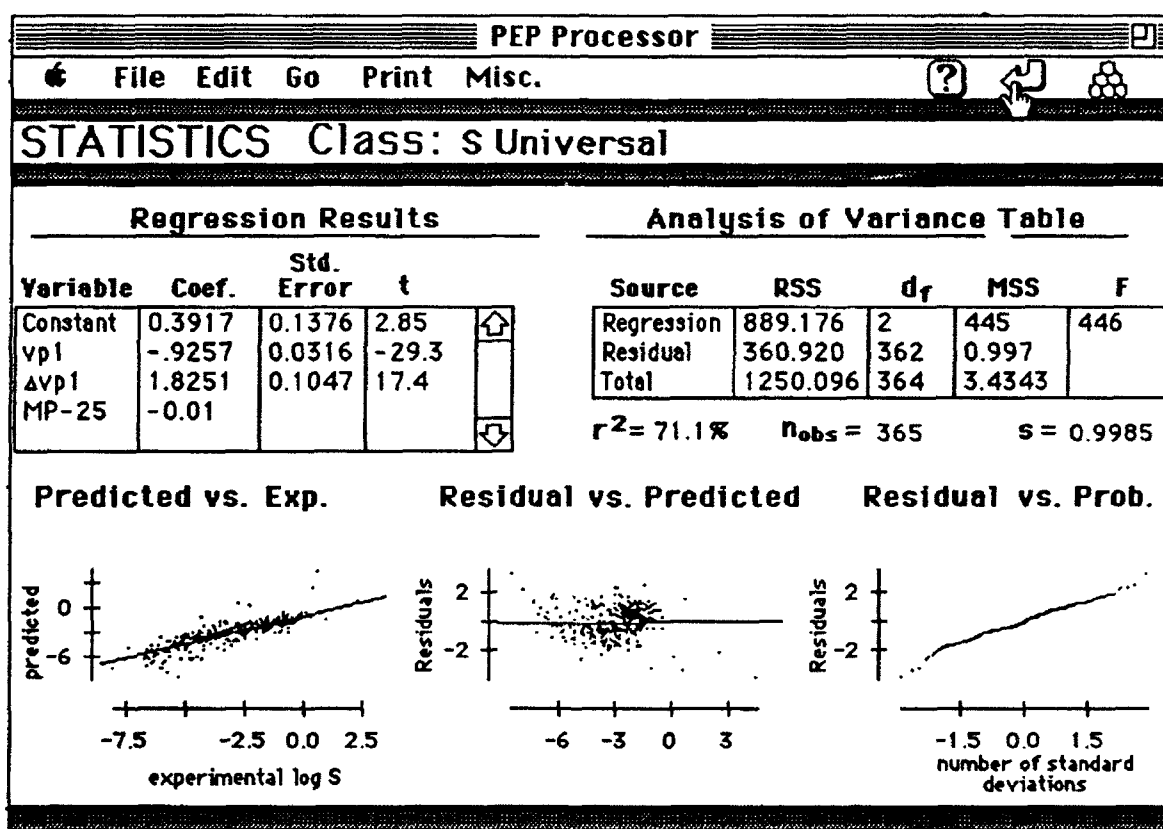


Figure 4. Example statistics card from PEP

The 2nd category of MCI-property correlations, located below the PEP relationships, were compiled from various literature sources. Clicking on the "book" icon will display the reference and information regarding the number and type of compounds included in the correlation if it was available in the original literature.

After choosing the most appropriate regression, estimates for the selected properties can be made. As shown in Figure 5, the MCI module results card provides an estimate of the property along with its calculated accuracy based on both the 95% confidence interval calculated from the regression and the estimated precision associated with the experimental determination of the property. In addition, the user has the option to search the property database for actual experimental values if they are available for comparison.

PEP Processor					
File Edit Go Print Misc. Models					
Estimated Properties					
Chemical Name: 2,2',6,6'-tetrachlorobiphenyl					
Method: MCI					
Property	*Value	Units	Regression Equation Used	View stats	Look for values in Prop.Db
log antilog					
log S	-6.17 ± 0.38	moles/L	PEP: PCBs		NA
log Kow	5.42 ± 0.43		PEP: PCBs		5.936
log Pv	-2.34 ± 1.04	Pascals	PEP: PCBs		NA
log H	-1.30 ± 0.28	dimensionless	PEP: PCBs		-1.648 dimensionless
log Koc	3.92 ± 1.20	L/kg	PEP: PCBs		NA
log BCF	4.69 ± 0.43		PEP: PCBs		NA
<p><b>*Note: The values shown are estimated at 25°C ± the 95% prediction interval for the regression used.</b></p>					

Figure 5. Results card from PEP MCI module.

**TSA module** The TSA module is similar in operation to the MCI module. However, unlike molecular connectivity, the calculation of TSA requires information describing the geometry of the molecule in terms of its 3-D atomic coordinates. The TSA module, shown in Figure 6, accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by commercially

available, Macintosh compatible, molecular modeling software such as Alchemy III™ or Chem3D Plus™.

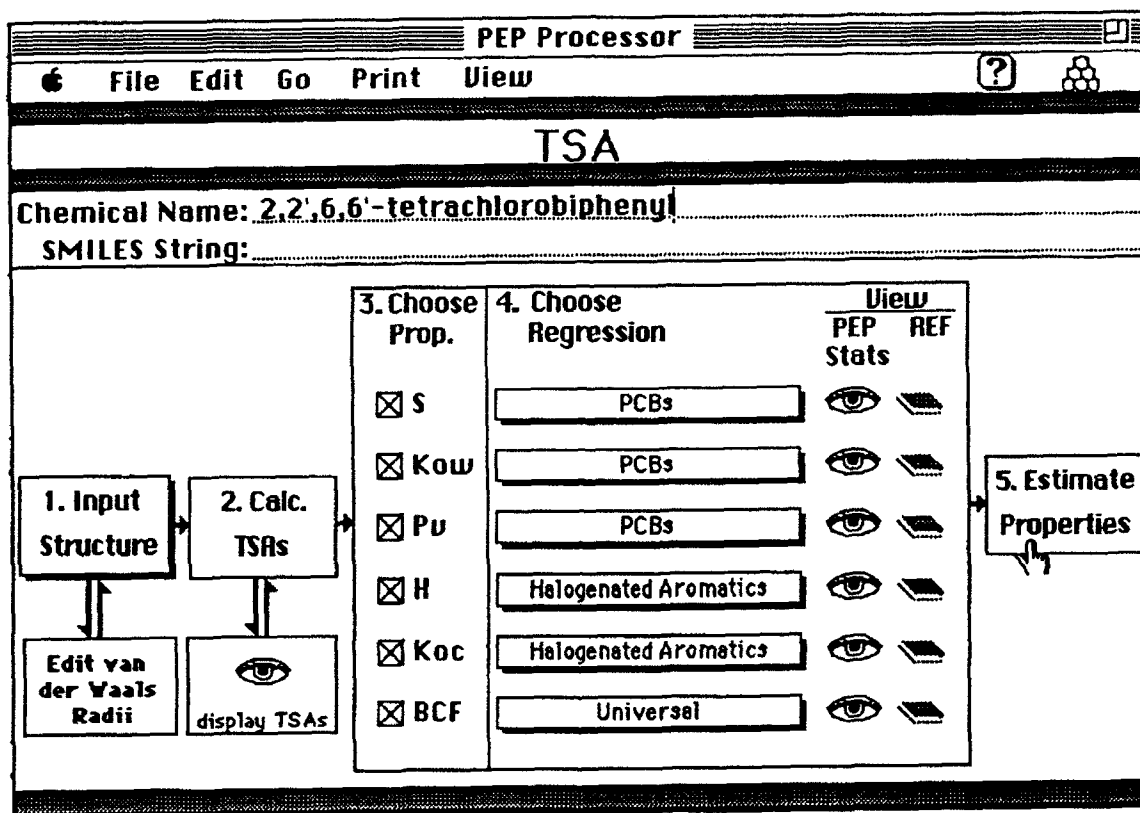


Figure 6. TSA module card from PEP

The TSA module is also designed to accept files generated by other hardware/software combinations including UNIX or VAX versions of CONCORD (Tripos Associates, Inc.), a hybrid expert system and molecular modeling software designed for the rapid generation of high quality approximate 3-D molecular structures. In addition to the 3-D molecular structure, the user must also input van der Waals radii for each of the atoms. A editable table of van der Waal radii, obtained from Bondi [37] for most common atoms, is provided within the TSA module. Once the molecular geometry and the van der Waal radii are input, TSA can be calculated using a XFCN which was adapted from the SALVO2 algorithm developed by Pearlman [19]. This algorithm



represents each atom of a molecule by a sphere centered at the equilibrium position of the nucleus. The radius of the sphere is equal to that of the van der Waals radius. Planes of intersection between spheres are used to estimate the contribution to surface area from the individual atoms or groups. The program computes the surface area of individual atoms or group by numerical integration, and the overlap due to intersecting spheres is excluded from the calculation. TSA is calculated by the summation of individual group contributions. The program also allows the TSA of the solute molecule to be calculated after the addition of a suitable solvent radius. A more detailed description of the TSA calculation method is provided by Pearlman [19].

After the TSA has been calculated, the user then chooses the properties of interest and a regression equation for each using the same approach as described in the MCI module. If the SMILES string or the connection table is also input, the most appropriate TSA-property relationship(s) will be flagged in the popup menu. The operation of the TSA module from this point on is identical to that of the MCI module.

UNIFAC module Like the MCI module, the UNIFAC module, illustrated in Figure 7, requires either a SMILES string or a connection table as input. An XFCN converts the structural information provided by the SMILES string or connection file into valid UNIFAC subgroups and counts the number of each subgroups present. In order to break the structure into the proper subgroups, the SMILES string or the connection file is interpreted and the information is put into a matrix. Each row and column in the matrix represents an atom in the chemical. The matrix contains the bond order between the two atoms that correspond to each entry in the matrix. If two atoms are not connected then a 0 is placed in the corresponding entry in the matrix. After the matrix is built the algorithm then "asks" specific questions about each atom, its neighbors, and how it is connected. If the answers to a set of questions are all true then a subgroup was found, the atoms are put together, and the matrix is reduced. The questions are then asked over again and the next subgroup is chosen, this repeats until no more subgroups are found. The questions are asked in a specific sequence so that the resulting subgroups are independent of the order of the atoms in the matrix.

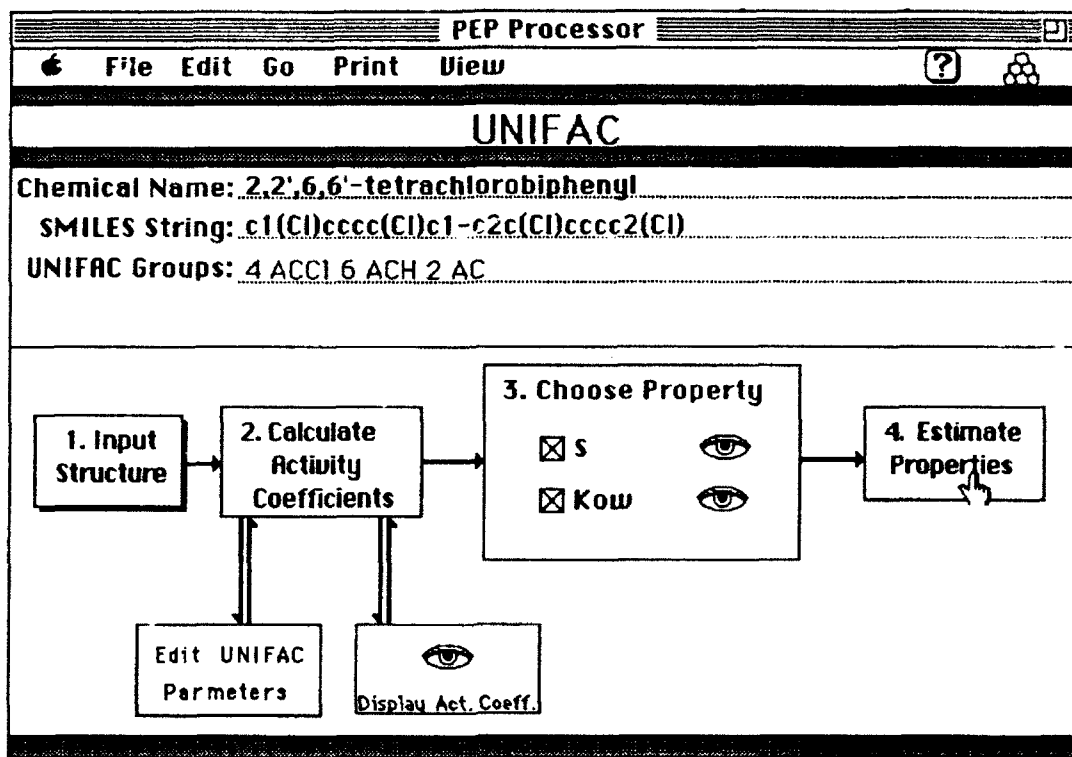


Figure 7. PEP UNIFAC module.

The UNIFAC method for calculating activity coefficients, as described by Grain [38], is implemented using both HyperTalk and an XFCN. The functional group interaction parameters, presented by Gmehling et al. [26] and derived from vapor-liquid equilibria (VLE), are used in the calculation routine but can be changed by the user. After the activity coefficients are calculated they can be displayed along with relevant intermediate values and used to estimate  $S$  and  $Kow$  by the following expressions (Arbuckle, 1986):

$$Kow = 0.115 \gamma_{ow} / \gamma_{oo} \quad (2)$$

$$S \text{ (mol/L)} = 55.6 / \gamma_{ow} \quad (3)$$

where  $\gamma_{ow}$  is the activity coefficient of the chemical infinitely dilute in water and  $\gamma_{oo}$  is the activity coefficient of the chemical infinitely dilute in octanol [27].

**Property/Property Module** Input for the Property/Property module, shown in Figure 8, depends on the the properties to be estimated and the regression models used. Thus, the user must

select the properties to be estimated and the property-property relationships (regression equations) to be used before any input values are requested. The program keeps track of the inputs required and provides the appropriate input fields. If available, the required properties can be imported directly from the associated chemical property database. Information regarding the regression statistics, if available, is also provided as previously described in the MCI module. After the necessary properties are entered into the corresponding input fields, the properties of interest can be estimated and the results, along with the 95% prediction interval (if the necessary data is available) can be viewed.

**PEP Processor**

File Edit Go Print View

**Property/Property**

Chemical Name: 2,2',6,6'-tetrachlorobiphenyl

SMILES String: \_\_\_\_\_

**1. Choose Property**

☒ S

☐ Kow

☒ Koc

☒ BCF

☐ H

**2.**

☒ USER choose regression

☐ PEP choose regression

	View	PEP	REF
	Stats		
Universal from Kow			
Universal from G & H			
Universal from Kow			
Universal from Kow			
P-LOG-KT			

**3. Input Properties**

Look in Prop. DB

log S \_\_\_\_\_

log Kow 5.936

log Pw \_\_\_\_\_

log H \_\_\_\_\_

log Koc \_\_\_\_\_

log BCF \_\_\_\_\_

M.P. \_\_\_\_\_ °C

M.W. \_\_\_\_\_

**4. Estimate Properties**

Note: All values are at 25°C

Figure 8. PEP Property/Property correlation module.

## PEP Models

To illustrate the practical application of PEP, an additional stack called PEP Models was developed. This stack, which contains the algorithms for the Level 1 and 2 Fugacity Models [32], is linked directly to the the PEP Processor, but can also be used independently.

The Level 1 Fugacity Model considers a unit world consisting of six compartments: air, water, soil, suspended solids, sediment, and biota as illustrated below in Figure 9. The model predicts the equilibrium concentrations of the chemical of interest in each compartment using the fugacity approach described by Mackay [32, 39]. The model requires the input of  $K_{OC}$ ,  $H$  and  $BCF$  which can be read directly from the PEP processor or the PEP chemical property database, if available. In addition to the chemical specific properties, the density and volume of each compartment must be specified along with the organic carbon content of the soil, sediment and suspended sediment. An editable set of default values for compartment density, volume and organic carbon content, as suggested by Mackay, is provided.

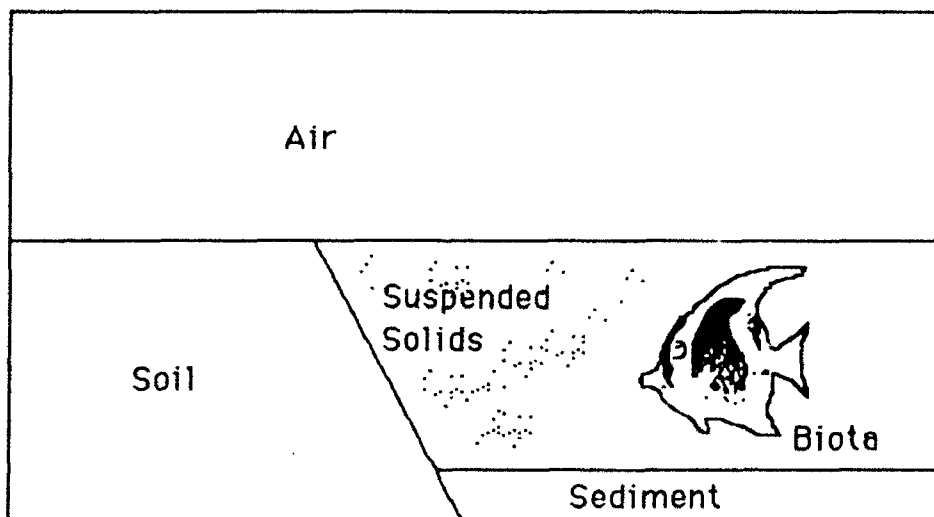


Figure 9. Representation of Fugacity Level 1 Model compartments.

The Level 2 Fugacity Model (Figure 10) allows for the chemical of interest to degrade in each compartment, move by advection through the water and air phases, and be emitted into the unit world. The rate values for each of these processes must be entered by the user. The degradation rates for each compartment can be entered either by  $t_{1/2}$  values in hours or by first order reaction rate constants in 1/hours. The advection rate data can be entered either by residence time or flow rate and the concentration or by directly entering the mass flow rate in moles per hour. The emission rate is entered in the units of moles per hour.

**PEP Models**

File Edit Go Print Compartments

### Fugacity Model

Chemical Name: 2,2',6,6'-tetrachlorobiphenyl

**1. Input Property Values**

Look for Values in Prop.DB

log Koc ☐ 3.92

log H ☒ (dimensionless) -1.648

log BCF ☐ 4.69

☒ Indicates value from DB

**2. Input Environmental Compartment Values**

Compartment	Density kg/m <sup>3</sup>	Volume m <sup>3</sup>	% Organic Carbon
<input checked="" type="checkbox"/> Air	1.19	1e10	
<input checked="" type="checkbox"/> Water	1000	7e6	
<input checked="" type="checkbox"/> Soil	1500	9e3	2
<input checked="" type="checkbox"/> Susp. Solids	1500	35	4
<input checked="" type="checkbox"/> Sediment	1500	2.1e4	4
<input checked="" type="checkbox"/> Biota	1000	3.5	

**3. Calculate Distribution**

OR

Input Fugacity Level 2 Data

Figure 10. PEP Models card

After the user inputs all the required information and hits the "calculate distribution" button, the model calculations are performed in HyperTalk and the results are presented in both in tabular and graphical form as illustrated in Figure 11. The graphical display can be changed from bar

(concentration of the chemical in each phase) to pie (percent of the chemical in each phase) chart forms using the "Graph:" popup menu. The values of the distribution coefficients that were used in the calculations are also shown on the results card. A complete description of these models has been given by Mackay [32,39]

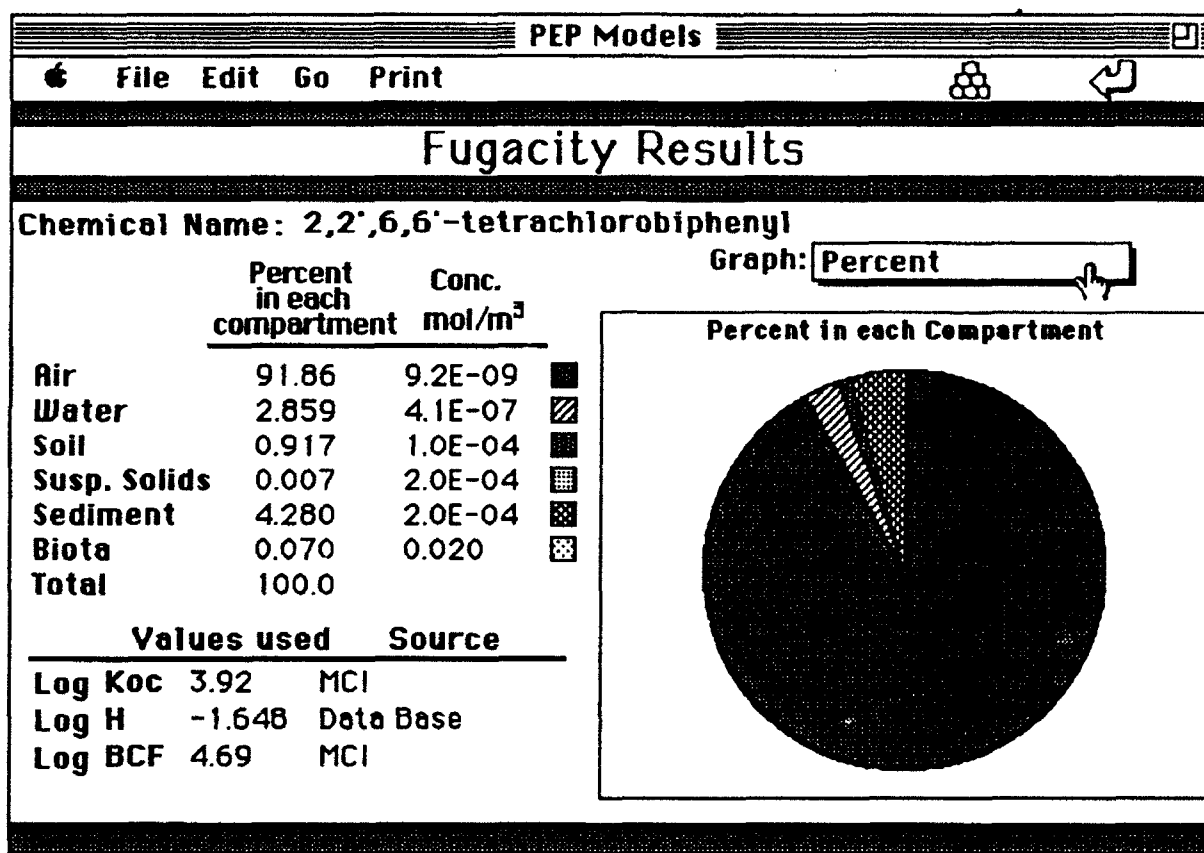


Figure 11. PEP Models results card

### PEP Help

Information regarding the operation of the chemical property database and the property estimation, models and batch modules is available in the PEP Help stack. This stack easily accessed at any time within the PEP system. The organization and layout of each help card is similar to that illustrated in Figure 12 for the MCI module. The user can select the topic of interest

by clicking on the appropriate radio button and the information on that subject will be displayed in the scrolling field.

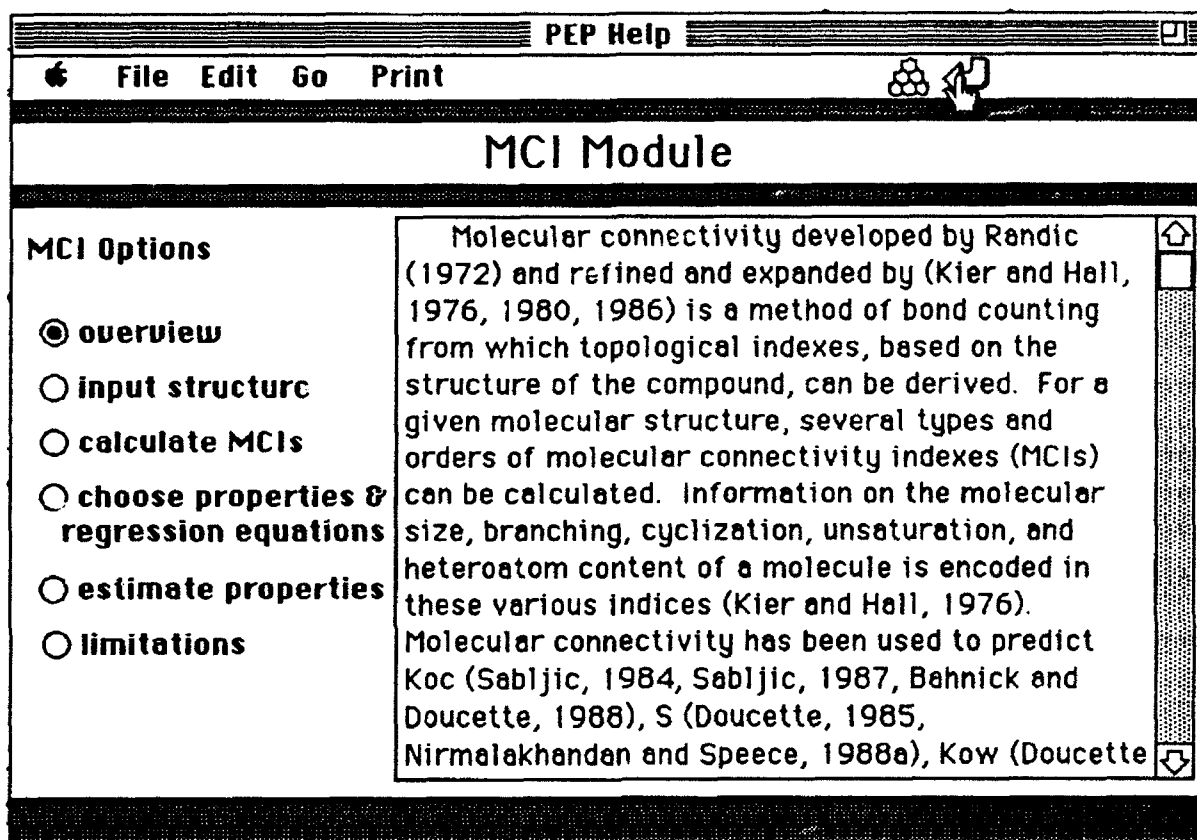


Figure 12. Example Help card from PEP

### Chemical Property Database

Experimentally determined physical property data for about 800 compounds, having at least one value of aqueous solubility (S), octanol/water partition coefficient (Kow), vapor pressure (Pv), organic carbon normalized soil sorption coefficient (Koc), bioconcentration factor (BCF), or Henry's law constant (H), was compiled from a variety of literature sources and computerized databases. Using this information, a chemical property database was constructed using HyperCard™ and subsequently used for developing MCI-property, TSA-property and property-property relationships. In addition to the properties listed above, the database includes the

following information: compound name and synonyms, a diagram of the 2-D chemical structure, SMILES notation, uses, CAS number, chemical formula, molecular weight (MW), boiling point (BP), melting point (MP), and appropriate references for each value. A built-in unit conversion utility enables users to quickly view property values in a variety of commonly used units. The database is directly connected to the PEP Processor stack.

The Chemical Property Database also provides the means for the user to search for chemical compounds by full or partial name or synonym, to sort the compounds by name, boiling point, melting point, or molecular weight, and the ability to transfer to any of the property estimation modules. In addition, the user can easily edit existing values, add new values or export information to a text file or another database.

In addition to the physical properties, information describing the environmental persistence and toxicity of specific chemicals can also be entered into the database. Placeholders for biodegradation rates, hydrolysis rates, photolysis rates and LC50s, along with the appropriate references and comments have been incorporated into the database. This feature was added to the database after requests from test users, however at the time of this report no degradation or toxicity data has been entered into the database. The chemical property database is illustrated in Figures 13 and 14.

### **PEP Batch**

PEP Batch provides users with a method for the convenient, unattended calculation of MCIs, TSA and UNIFAC activity coefficients and subsequent estimation of physical properties for large numbers of compounds via the PEP processor described earlier. Like the PEP Processor, PEP Batch is divided into MCI, TSA and UNIFAC modules. Each module, as illustrated in Figure 15 for the MCI module, requires the user to select the appropriate input file (i.e. SMILES string, connection table or 3-D atomic coordinates), choose to information to be sent to the output file (i.e. chemical name, SMILES string, properties) and start the batch driver.



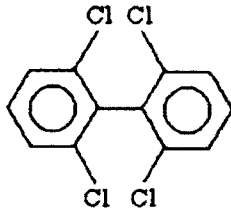
Chemical Property Data Base				
File Edit Go Find Sort Values Print Export		179		
<b>2,2',6,6'-tetrachlorobiphenyl</b>		View: <b>Structure</b>		
CAS #: 15968-05-5				
Formula: _____ MW: 292				
MP: 200 — °C BP: _____ °C				
Synonyms: _____				
View: <b>Chemical\Physical Properties</b>				
Antilog	Physical/Chemical Properties	Values	Temp °C Units Ref	
log	Aqueous Solubility (S)			
log	Octanol/Water Partition Coefficients (Kow)	5.936	25	
log	Vapor Pressure (Pv)			
log	Henry's Law Constant (H)	-1.648	25	dimensionless
log	Soil/Water Sorption Coefficient (oc-based) (Koc)			
log	Bioconcentration Factor (BCF)			
log	Acid Dissociation (Ka)			

Figure 13 Example card from the Chemical Property Data Base

chemical property data base

File Edit Go Find Sort Values Print Export

2,2',6,6'-tetrachlorobiphenyl

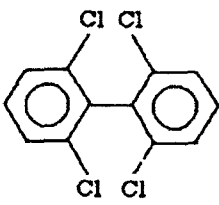
CAS #: 15968-05-5

Formula: MW: 292

MP: 200 °C BP: °C

Synonyms:

View: Structure



View: Chemical\Physical Properties

Degradation Properties				Comments
Degradation	t 1/2			
Hydrolysis				
	t 1/2	Matrix	Temp (°C)	Comments
Photolysis				
	t 1/2	Matrix	Temp (°C)	Comments
Biodegradation				

Figure 14. PEP database degradation properties.

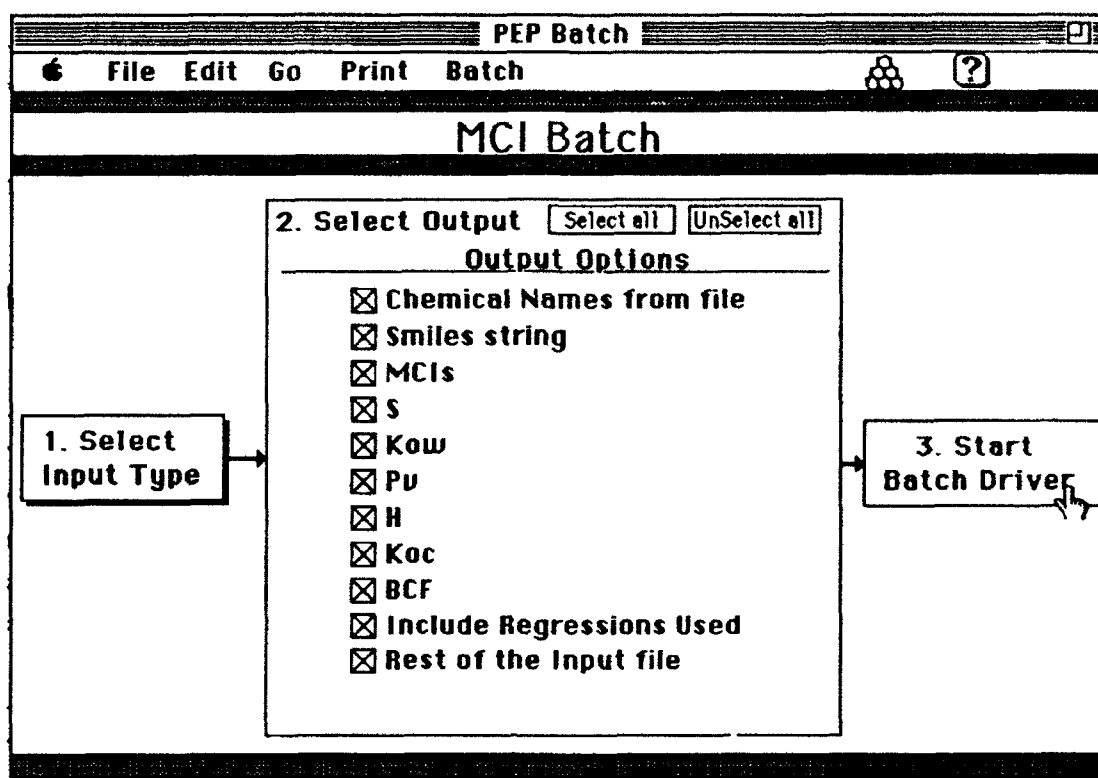


Figure 15. Example card for PEP Batch, MCI module.

The MCI and UNIFAC modules require the two-dimensional molecular structure of the chemical to be entered using either SMILES strings or connection tables.

To enter SMILES strings into the MCI or UNIFAC batch modules, you must create a text file containing name of the chemical and its corresponding SMILES strings in columns separated by tabs and hard returns at the end of each row. The text file can contain additional tab-delimited information, but the SMILES strings and chemical names must be in the first or second column. You indicate the column order when you select the type of input. The text file can be created with word processing or spreadsheet programs or you can also edit or create a file containing the SMILES strings and chemical names within PEP by selecting SMILES from the "input structure type" popup menu.

To input connection table files into the MCI or UNIFAC batch modules, you must first place them in a single folder. Select the "Connection tables" option from the "Select Input Type" popup

button and choose the folder that contains the connection table files using the standard "open file" dialog box that appears. Highlighting any one of the files in the folder selects all of the files in that folder and allows you to view or delete specific files. The files can be any valid type for the MCI module as they will be converted if possible (see MCI module helps). After selecting the input files, click on the advance arrow located at the upper right of the card to advance to the next step.

The TSA batch module operates in the same manner as the MCI and UNIFAC modules except that the calculation of TSA requires the three-dimensional cartesian coordinates for each atom in the chemical of interest. The TSA batch module accepts Cartesian Coordinates or Alchemy files. Alchemy files contain both the two-dimensional chemical structure and the coordinates. This allows PEP to calculate the chemical's TSA and determine the most appropriate TSA-property relationship based on chemical class. From within the standard dialog box, you can click on any file in a folder to select all of the files in that folder. The files will then be displayed. You can also view or delete files.

Once the input files have been selected, the "output option" step becomes active. This allows the user to select the properties to be calculated and any additional information that is available (i.e. chemical name, SMILES string, MCIs, TSA, or UNIFAC activity coefficients) to be exported by to a tab-delimited text file.

After the output information is selected, the "Start Batch Driver" button becomes active. Clicking this button brings up the standard Macintosh "save file" dialog box that allows the user to specify the name of the data file to be exported by PEP batch and the location that the file will be sent.

## SUMMARY

A microcomputer program for estimating physical/chemical properties of organic chemicals for use in environmental fate modeling has been described. The **Property Estimation Program (PEP)** and associated physical property database was developed using HyperCard for the Apple Macintosh series of computers. The PEP system utilizes both QSPRs and QPPRs to provide the user with several approaches to estimate  $S$ ,  $K_{ow}$ ,  $P_v$ ,  $H$ ,  $K_{oc}$  and  $BCF$  depending on the information available. While QPPRs have been used by both experts and non-experts for estimating properties, one of the major limitations in using QSPRs has been the difficulty in using the necessary software tools. The graphical interface and flow chart design of PEP leads the user through a series of logical steps designed to provide even non-experts with a economical, easy to use software system for property estimation. The structural information for the MCI and UNIFAC modules can be input using Simplified Molecular Input Line Entry System (SMILES) notation or connection tables generated from a commercially available two-dimensional drawing program. The TSA module accepts 3-D cartesian coordinates entered manually or directly reads coordinate files generated by molecular modeling software. For each property the user can select from either "universal" or class specific regression models. The program's built in intelligence helps the user choose the most appropriate QSPR based on the structure of the chemical of interest. In addition, sufficient statistical information is provided to allow the user to determine on the validity of the QSPRs and QPPRs utilized in PEP. Designed to make the program both practical and educational, on line documentation is provided not only for the operational characteristics of the program but also for the theory associated with the property estimation techniques.

The combination of the various property estimation methods, chemical property database, and simple environmental fate models provides users with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input.

## MISCELLANEOUS PUBLICATIONS

- Frazier, J.D. 1990. Estimation of Chemical/Physical Properties Using Molecular Connectivity Indices for Application in Modeling the Environmental Fate of Organic Compounds. M.S.thesis, Utah State University.
- Holt, M. S. 1992. Microcomputer Program for the Estimation of Properties for Use in Environmental Fate and Transport Models. M.S.thesis, Utah State University.
- Denne, D. 1993 (expected completion, March 1993). Investigating the Relation Between Molecular Structure and Parameters Important in the Environmental Fate of Organic Chemicals. M.S.thesis Utah State University.
- Miles, R. 1993 (expected completion, May 1993). A simple and inexpensive method for measuring biodegradation rates in soil. M.S.thesis, Utah State University

## LIST OF PAPERS/POSTERS PRESENTED AT PROFESSIONAL MEETINGS

- Miles, R., W.J. Doucette, B. Bugbee. 1992. A simple and inexpensive method for measuring biodegradation rates in soil. Presented in a poster session at the 13th Annual SETAC meeting, November 8-12, Cincinnati, OH.
- Li, A., W.J. Doucette, A.W. Andren. 1992. UNIFAC estimation of environmental parameters for PCBs. Presented in a platform session at the 13th Annual SETAC meeting, November 8-12, Cincinnati, OH.
- Doucette, W. J. and M. S. Holt. 1992. PEP, a microcomputer program for estimating physical/chemical properties. Presented as a poster at the Fifth International Workshop on QSAR in Environmental Toxicology (QSAR 92), Duluth, MN (July 19-23):
- Holt, M.S. and W.J. Doucette. 1992. A Hypercard Based Program for the Estimation of Physical/Chemical Properties Used in Environmental Fate Modeling. Presented as a platform presentation at the First Annual Conference on Scientific and Engineering Applications of the Macintosh (SEAM'92), San Francisco, CA: 1992 (January 15-17).
- W. J. Doucette, D.J. Denne, M. S. Holt and J. E. McLean. 1991. Prediction of Physical/Chemical Properties Using MCI and TSA-Property Relationships. Presented in platform session at the 12th Annual SETAC meeting, November 3-7, Seattle, WA
- Holt, M.S. and W.J. Doucette. 1991. Microcomputer program utilizing Quantitative Structure Property Relationships (QSPRs) and Quantitative Property Property Relationships (QSPRs)

for the Estimation of physical/chemical properties used in environmental fate modeling  
Presented in a platform session at the 9th National Conference on Microcomputers in Civil  
Engineering, October 31, Orlando, Florida

Doucette, W. J., D.J. Denne, J.D. Frazier, R.R. Dupont, D.K. Stevens, M. E Holt, and J.E.  
McLean. 1990. Microcomputer program utilizing quantitative structure-property and  
property-property relationships for environmental fate modeling. Presented in poster session at  
11th Annual SETAC meeting, November 11-15, Arlington, VA

Doucette, W.J., D.J. Denne, J.D. Frazier, R.R. Dupont, D.K. Stevens, M. E Holt, and J.E.  
McLean. 1990. Microcomputer program utilizing molecular connectivity indices, UNIFAC-  
derived activity coefficients, and property-property correlations for the estimation of physical  
properties. Presented in poster session at the 4th International Workshop on QSARs in  
Environmental Toxicology, September 16-20, Veldhoven, The Netherlands .

Doucette, W.J., H. Fugate, J.D. Frazier, and D.A. Bahnick. 1989. Structure Property  
Relationships Utilizing Molecular Connectivity Indices and Total Molecular Surface Areas for  
Environmental Fate Modeling. Presented at the International Chemical Congress of Pacific  
Basin Societies, December 21, Honolulu, Hawaii, USA

Frazier, J.D, and W.J. Doucette. 1989. Microcomputer Program Utilizing Molecular Connectivity  
Indices for Property Estimation. Presented at the International Chemical Congress of Pacific  
Basin Societies, December 21, Honolulu, Hawaii, USA

### **LIST OF GRADUATE STUDENTS ASSOCIATED WITH THE RESEARCH EFFORT**

Joe Frazier, M.S. 1990 Thesis title: Estimation of Chemical/Physical Properties Using Molecular  
Connectivity Indices for Application in Modeling the Environmental Fate of Organic  
Compounds.

Mark Holt, M.S. 1992 Thesis title: Microcomputer Program For The Estimation Of Properties  
For Use In Environmental Fate And Transport Models

Denne, D. J. expected March 1993. Tentative thesis title: Investigating the Relation Between  
Molecular Structure and Parameters Important in the Environmental Fate of Organic Chemicals.  
Utah State University.

Rick Miles, M.S. expected May 1993. Tentative thesis title: A simple and inexpensive method  
for measuring biodegradation rates in soil.

## REFERENCES

1. Kenaga, E.E. and C.A.I. Goring, Relationship Between Water Solubility, Soil Sorption, Octanol-Water Partitioning, and Concentration of Chemicals in Biota. 1980. American Society for Testing and Materials: p. 78-115.
2. Karickhoff, S.W., Semi-Empirical Estimation of Sorption of Hydrophobic Pollutants on Natural Sediments and Soils. *Chemosphere*, 1981. 10(8): p. 833-846.
3. Mackay, D., Correlation of Bioconcentration Factors. *Environ. Sci. Technol.*, 1982. 16(5): p. 274-278.
4. Randic, M., On Characterization of Molecular Branching. *J. Amer. Chem. Soc.*, 1972. 97(23): p. 6609-6615.
5. Kier, L.B. and L.H. Hall, Molecular connectivity in chemistry and drug research. in *Medicinal Chemistry a Series of Monographs*. 1976, N. Y.: DeSteven, G. ed. Academic Press.
6. Kier, L.B., Molecular connectivity as a description of structure for SAR analysis in *Physical Chemical Properties of Drugs*. 1980, New York, NY: Yalkowsky, S. H., A. A. Sikula, and S. C. Valvani, eds. Marcel Dekker, Inc. 331pp.
7. Kier, L.B., Hall, and L. H., *Molecular Connectivity in Structure-Activity Analysis*. 1986, Letchworth, Hertfordshire, England.: Research Studies Press, LTD., England, and John Wiley & Sons, Inc., New York. 262.
8. Sabljic, A., Predictions of the Nature and Strength of Soil Sorption of Organic Pollutants by Molecular Topology. *J. Agric. Food Chem.*, 1984. 32: p. 243-246.
9. Sabljic, A., On the Prediction of Soil Sorption Coefficients of Organic Pollutants from Molecular Structure: Application of Molecular Topology Model. *Environ. Sci. Technol.*, 1987. 21: p. 358-366.
10. Doucette, W.J. and A.W. Andren, Estimation of Octanol/Water Partition Coefficients: Evaluation of Six Methods for Highly Hydrophobic Aromatic Hydrocarbons. *Chemosphere*, 1988. 17(2): p. 334-359.
11. Nirmalakhandan, N.N. and R.E. Speece, QSAR Model for Predicting Henry's Constant. *Environ. Sci. Technol.*, 1988. 22(11): p. 1349-1357.



12. Briggs, G.G., Theoretical and Experimental Relationships between Soil Adsorption, Octanol-Water Partition Coefficients, Water Solubilities, Bioconcentration Factors and the Parachor. *J. Agric. Food Chem.*, 1981. 29: p. 1050-1059.
13. Hermann, R.B., Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.*, 1972. 76: p. 2754-2759.
14. Amidon, G.L., S.H. Yalkowsky, and S.Leung, Solubility of Nonelectrolytes in Polar Solvents II. Solubility of Aliphatic Alcohols in Water. *J. Pharm. Sci.*, 1974. Vol. 63(No. 12): p. 1858-1866.
15. Yalkowsky, S.H. and S.C. Valvani, Partition Coefficients and Surface Areas of Some Alkylbenzenes. *Journal of Medicinal Chemistry*, 1976. Vol. 19, No. 5: p. 727-728.
16. Yalkowsky, S.H. and S.C. Valvani, Solubilities and Partitioning 2. Relationships Between Aqueous Solubilities, Partition Coefficients, and Molecular Surface Areas of Rigid Aromatic Hydrocarbons. *Journal of Chem.*, 1979. 24(2): p. 127-129.
17. Mackay, D., R. Mascarenhas, and W.Y. Shiu, Aqueous Solubility of Polychlorinated Biphenyls. *Chemosphere*, 1980. 9: p. 257-264.
18. Amidon, G.L. and S.T. Anik, Application of the Surface Area Approach to the Correlation and Estimation of Aqueous Solubility and Vapor Pressure. Alkyl Aromatic Hydrocarbons. *Journal of Chemical and Engineering Data*, 1981. Vol. 26 No. 1: p. 28-33.
19. Pearlman, R.S., Molecular Surface Areas and Volumes and Their Use in Structure/Activity Relationships, in *Physical Chemical Properties of Drugs*, S.H. Yalkowsky, A.A. Sikula, and S.C. Valvani, Editor. 1980, Marcel Dekker, Inc.: New York. p. 331-346.
20. Doucette, W.J. and A.W. Andren, Correlation of Octanol/Water Partition Coefficients and Total Molecular Surface Area of Highly Hydrophobic Aromatic Compounds. *Environ. Sci. Technol.*, 1987. 21: p. 821-824.
21. Hawker, D.W. and D.W. Connell, Octanol-Water Partition Coefficients of Polychlorinated Biphenyl Congeners. *Environ. Sci. Technol.*, 1988. 22(4): p. 382-387.
22. Fugate, H.N., Using Total Molecular Surface Area in Quantitative Structure Activity Relationships to Estimate Environmental Fate and Transport Parameters. 1989, Utah State University:
23. Connolly, M.L., Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science*, 1983. 221(4612): p. 709-713.

24. Fredenslund, A., R.L. Jones, and J.M. Prausnitz, Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE Journal*, 1975. 21(6): p. 1086-1099.
25. Fredenslund, A.J., Gmehling, and P. Rasmussen, Vapor-liquid equilibria using UNIFAC, a group-contribution method. 1977, New York, NY 213 pp.: Elsevier Scientific Publishing Co.
26. Gmehling, J., P. Rasmussen, and A. Fredenslund, Vapor-liquid equilibria by group contribution. Revision and Extension. 2. *Ind. Eng. Chem. Process Des. Dev.*, 1982. 21: p. 118-127.
27. Arbuckle, W.B., Estimating Activity Coefficients for Use in Calculating Environmental Parameters. *Environ. Sci. Technol.*, 1983. 17: p. 537-542.
28. Banerjee, S., Calculation of Water Solubility of Organic Compounds with UNIFAC-Derived Parameters. *Environ. Sci. Technol.*, 1985. 19(4): p. 369-370.
29. Arbuckle, W.B., Using UNIFAC to Calculate Aqueous Solubilities. *Environ. Sci. Technol.*, 1986. 20: p. 1060-1064.
30. Doucette, W.J., Measurement and Estimation of Octanol/Water Partition Coefficients and Aqueous Solubilities for Halogenated Aromatic Hydrocarbons. 1985, PhD. thesis University of Wisconsin:
31. Banerjee, S. and P.H. Howard, Improved Estimation of solubility and Partitioning Through Correction of UNIFAC-Derived Activity Coefficients. *Environ. Sci. Technol.*, 1988. 22(7): p. 839-841.
32. Mackay, D., Finding Fugacity Feasible. *Environ. Sci. and Technol.*, 1979. 13(12): p. 1218-1223.
33. Anderson, E., G.D. Veith, and D. Wieninger, SMILES: A Line Notation and Computerized Interpreter for Chemical Structures. 1987. EPA/600/M-87/021.
34. Weininger, D., SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.*, 1988. 28: p. 31-36.
35. Frazier, J.D., Physical/Chemical Property Estimation Using Molecular Connectivity Indices for Application in Environmental Fate Modeling. 1990, Utah State University:
36. Bahnick, D.A. and W.J. Doucette, Use of Molecular Connectivity Indices to Estimate Soil Sorption Coefficients for Organic Chemicals. *Chemosphere*, 1988. 17(9): p. 1703-1715.

37. Bondi, A., Van der Waals Volumes and Radii. *Journal of Physical Chemistry*, 1964. Vol 68(Number 3): p. 441-451.
38. Grain, Activity Coefficients, in *Handbook of Chemical and Property Estimation Methods*, W.J. Lyman, W.F. Reehl, and D.H. Rosenblatt, Editor. 1982, American Chemical Society: Washington D. C.
39. Mackay, D. and S. Paterson, Calculating Fugacity. *Environ. Sci. Technol.*, 1981. 15(9): p. 1006-1014.
40. Hodgdon, Jay. 1988a. ShowDialog. Ignatz Software. Foster City, California [Computer Program]
41. Hodgdon, Jay. 1988b. Progress. Ignatz Software. Foster City, California [Computer Program]
42. Freed, Adrain. 1989a. XTRA Shell. Fidcor USA. Louisville, Colorado. [Computer Program]
43. Freed, Adrain. 1989b. PopUp. Fidcor USA. Louisville, Colorado. [Computer Program]
44. Symantec Corporation. 1991. Think C vs 4.0 and 5.0. Cupertino, California. [Computer Program]
45. ChemDraw 1989. Cambridge, Massachusetts [Computer Program]
46. Cobalt Blue. 1989. FOR\_C, A FORTRAN to C translator. San Jose California. [Computer Program]
47. Abacus Concepts Inc. 1988. StatView II. Berkeley, California. [Computer Program]
48. Odesta Corp. 1989. Data Desk. Northbrook, Illinois [Computer Program]

**APPENDIX A**

**PEP**  
**Property Estimation Program**  
**and**  
**Chemical Property Database**

**PEP**  
**Property Estimation Program**  
**and**  
**Chemical Property Database**

**Utah Water Research Laboratory**  
**Utah State University**  
**Logan, Utah 84321**

---

# TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . .	vii
DISCLAIMER . . . . .	viii
INTRODUCTION . . . . .	1
Background . . . . .	1
PEP Overview . . . . .	2
PEP Features . . . . .	3
What Do I Need to Use PEP? . . . . .	4
Installation of PEP . . . . .	4
General Programming Description . . . . .	5
Starting the PEP Software System . . . . .	5
Menus, Buttons and Icons . . . . .	7
Tutorial . . . . .	8
REFERENCE SECTION . . . . .	10
PEP Processor . . . . .	10
MCI Module . . . . .	11
Overview . . . . .	11
Entering Chemical Structure . . . . .	12
Calculating MCIs . . . . .	14
How PEP Calculates MCIs . . . . .	15
Choosing the Properties . . . . .	16
Choosing the Regression Models and Chemical Classes . . . . .	16
Statistics Cards . . . . .	18
Estimating the Properties and Viewing the Results . . . . .	19
Adding or Deleting MCI-Property Regression Models . . . . .	19
Limitations of MCI-Property Regression Models . . . . .	20
TSA Module . . . . .	22
Overview . . . . .	22
Entering the Structural Information . . . . .	23
Calculating Total Surface Area (TSA) . . . . .	25
Choosing the Most Appropriate TSA-Property Regression Model . . . . .	27
Estimating the Properties . . . . .	28
Development of TSA-Property Relationships . . . . .	28
Limitations of TSA-Property Relationships . . . . .	29

---

## TABLE OF CONTENTS (Cont'd)

UNIFAC Module . . . . .	30
Overview . . . . .	30
Entering Structural Information . . . . .	31
Calculate Activity Coefficients . . . . .	34
Editing Parameters . . . . .	34
Estimating Properties . . . . .	34
Limitations of UNIFAC Approach to Estimating S and $K_{ow}$ . . . . .	34
Property/Property Module . . . . .	35
Overview . . . . .	35
Selecting the Properties to be Estimated . . . . .	36
Choosing the Property-Property Regression Model . . . . .	36
Viewing the Calculated Values . . . . .	36
Limitations of the Property/Property Module . . . . .	37
PEP Batch . . . . .	38
Overview . . . . .	38
Input Structure . . . . .	39
Output Options . . . . .	40
Start Batch Driver . . . . .	40
CHEMICAL PROPERTY DATABASE . . . . .	41
Overview . . . . .	41
Searching for Chemical Compounds in PEP's Database . . . . .	43
Sorting the Database . . . . .	43
Adding or Deleting Data . . . . .	43
Printing Information from the Database . . . . .	44
Exporting Information from the Database . . . . .	44
Moving to Other PEP Modules . . . . .	44
Changing the Units of Measurement . . . . .	44
PEP Models . . . . .	45
Overview . . . . .	45
Input Property Values . . . . .	47
Input Environmental Values . . . . .	48
Calculate Distribution . . . . .	48
PEP Help . . . . .	50
REFERENCES . . . . .	51

---

## LIST OF TABLES

	<u>Page</u>
1 MCI's Calculated by PEP . . . . .	15
2. UNIFAC Groups . . . . .	33



---

## LIST OF FIGURES

	<u>Page</u>
1. PEP stack icons . . . . .	1
2. PEP opening screen . . . . .	6
3. The steps for using a menu for command selection . . . . .	7
4. Buttons and icons used in PEP . . . . .	8
5. Opening screen of PEP tutorial . . . . .	9
6. Example illustrating the use of the PEP Processor's flow chart interface . . . . .	10
7. Screen display of PEP MCI module . . . . .	11
8. Standard Macintosh file selection dialog box . . . . .	12
9. Example connection table . . . . .	13
10. Statistics card associated with the MCI module . . . . .	18
11. Example dialog box for the input of new MCI-property relationships . . . . .	20
12. TSA module card from PEP . . . . .	22
13. Example Alchemy file . . . . .	24
14. Example Cartesian coordinate file . . . . .	24
15. Example card for the entry of atomic coordinates . . . . .	25
16. Dialog box for editing van der Waal radii . . . . .	26
17. Example TSA card . . . . .	27
18. Example card from the PEP UNIFAC module . . . . .	31
19. UNIFAC module card used to select UNIFAC groups . . . . .	32
20. Example card from PEP property-property module . . . . .	35
21. Example card for PEP Batch, MCI module . . . . .	38

---

## LIST OF FIGURES (Cont'd)

	<u>Page</u>
22. PEP Batch, explain SMILES file card . . . . .	39
23. Example card from the Chemical Property data base . . . . .	42
24. PEP database degradation properties . . . . .	42
25. Representation of Fugacity Level 1 compartments . . . . .	45
26. PEP Models card, input for Fugacity Level 1 . . . . .	46
27. PEP Models card, input for Fugacity Level 2 . . . . .	47
28. Example dialog box resulting rom the "Look for Values in Prop.DB" option . . . . .	48
29. PEP Models results card . . . . .	49
30. Example card from PEP Help stack for MCI module . . . . .	50

---

## **ACKNOWLEDGMENTS**

PEP was initially developed with funding provided by the Air Force Office of Scientific Research, Bolling Air Force Base, DC 20332-6448, Grant No. AFOSR-89-0509

Project Leader: Lt. Col. T. Jan Cervený, Life Science Directorate, Department of the Air Force.

Authors: William J. Doucette and Mark S. Holt

Project Investigators: W. J. Doucette (Principal Investigator), D. K. Stevens, R. R. Dupont, R. C. Sims, and J. E. McLean

Graduate Students: Mark Holt, Doug Denne, Rick Miles and Joe Frazier

Programmers: Mark Holt, Joe Frazier and Mike Jablonski (NR Systems, Inc.)

The authors would also like to acknowledge the following individuals for their assistance with the SMILES interpreter: Eric Anderson, Gil Veith, Chris Russom (U.S. EPA, ERL-Duluth)

---

## ***DISCLAIMER***

Utah State University and the authors of PEP make no warranties, either expressed or implied with respect to the operation or subsequent use of property values obtained through the use of PEP. In no event will Utah State University or the authors of PEP be liable for direct, indirect, or consequential damages resulting from the use of this software.

---

## **INTRODUCTION**

### **Background**

Mathematical models are often used by environmental scientists and engineers to estimate the fate and impact of organic chemicals in the environment. Use of these models requires a variety of parameters describing site and chemical characteristics. Aqueous solubility (S), the octanol/water partition coefficient (K<sub>ow</sub>), the organic carbon normalized soil/water sorption coefficient (K<sub>oc</sub>), vapor pressure (P<sub>v</sub>), Henry's Law constant (H), and bioconcentration factor (BCF) are considered key properties used to assess the mobility and distribution of a organic chemical in environmental systems.

One major limitation to the use of environmental fate models has been the lack of suitable values for many of these properties. The scarcity of data, due mainly to the difficulty and cost involved in experimental determination of such properties for an increasing number of synthetic chemicals, has resulted in an increased reliance on the use of estimated values.

Quantitative Property-Property Relationships (QPPRs), based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property. Quantitative Structure-Property Relationships (QSPRs) often take the form of a correlation between a structurally derived parameter(s), such as molecular connectivity indices (MCI<sub>s</sub>) or total molecular surface area (TSA) and the property of interest.

Selection and application of the most appropriate QPPRs or QSPRs for a given compound is based on several factors including: the availability of required input, the methodology for calculating the necessary structural or topological information, the appropriateness of correlation to chemical of interest and an understanding of the mechanisms controlling the property being estimated.

Incorporation of QPPRs and QSPRs into a computer format is a logical and necessary step to gain full advantage of the methodologies for simplifying fate assessment.

## PEP Overview

A Property Estimation Program (PEP), utilizing MCI-property, TSA-property and property-property correlations and UNIFAC-derived activity coefficients, has been developed for the Apple Macintosh microcomputer to provide the user with several approaches to estimate S, Kow, Pv, H, Koc and BCF depending on the information available.

Structural information required for the MCI and UNIFAC calculation routines can be entered using either Simplified Molecular Identification and Line Entry System (SMILES) notation or connection tables generated with commercially available two-dimensional drawing programs. The TSA module accepts 3-D atomic coordinates entered manually or directly reads coordinate files generated by molecular modeling software. The program's built-in intelligence helps the user choose the most appropriate QSPR or QPPR based on the structure of the chemical of interest. In addition, the statistical information associated with each QSPR or QPPR in PEP can be displayed to help the user determine the model's validity. For the regression-based property estimation models, assessments of accuracy based on the 95% confidence interval and estimated precision of the experimental values are also provided along with the estimated property value.

PEP also provides a batch mode that provides users with a method for the convenient, unattended calculation of MCIs, TSA and UNIFAC activity coefficients and the subsequent estimation of physical properties for large numbers of compounds.

A chemical property database, containing experimental values of S, Kow, H, Pv, Koc, and BCF compiled from a variety of literature sources and computerized databases was used for developing the MCI-property, TSA-property and property-property relationships used in PEP. This database, which currently contains over 800 chemicals, is linked directly to PEP.

The property estimation modules in PEP are also linked directly to the Level 1 and 2 Fugacity Models. The combination of the various property estimation methods, chemical property database, and simple environmental fate models provides users with a methodology for predicting the environmental distribution of an organic chemical in a multi-phase system requiring only the structure of the chemical of interest as input.

PEP was designed to be intuitive and user friendly. The easiest way to become familiar with the PEP is to try clicking on the buttons and pull down menus found on each card. Any comments or suggestions regarding improving the operation of PEP would be greatly appreciated by the authors.

## • PEP Features

- Developed using Hypercard™ for the Apple Macintosh series of personal computers
- Comprised of a chemical property database and four property estimation modules
- Uses standard Macintosh operations (buttons, menus, windows)
- Simple user interface based on flow chart design
- Four property estimation methods are available:
  - Molecular Connectivity Indices (MCIs)-property correlations
  - Total Surface Area Regressions (TSA)-property correlations
  - Property-Property Correlations
  - UNIFAC derived activity coefficients
- PEP can be used to estimate six chemical/physical properties
  - Solubility (S)
  - Octanol-water partition coefficients (Kow)
  - Henry's Law Constant (H)
  - Vapor Pressure (Pv)
  - Organic carbon normalized soil-water distribution coefficients (Koc)
  - Bioconcentration factors (BCF)
- Universal and class specific regression models are available
- PEP uses decision support for determination of chemical class.
- Estimates include 95% prediction interval for each regression based estimated value
- Statistical information readily available for each regression
- New regression models can be easily added
- Database contains over 800 chemicals having at least one property values
- Each chemical has at least one property and a two-dimensional SMILES string
- Chemicals in database can be search for by chemical name, CAS number, synonym, or selected from an alphabetized list
- Property estimation modules and property database are linked directly to the Fugacity Level 1 and 2 environmental fate models
- Published and on-line documentation
- Includes PEP tutorial
- Includes PEP Batch for estimating properties or calculating MCIs, TSA, or UNIFAC activity coefficients for large numbers of chemicals without continuous user input

• **What Do I Need To Use PEP (i.e. Hardware requirements)?**

1. Macintosh II computer or better with
2. 4,000 Kbytes (4 Megabytes) of usable hard disk space,
3. 3 Meg of RAM installed,
4. running system software 6.0.5 or higher, and
5. HyperCard 2.0 software or higher installed with the size allocated to 1500 MB.

• **Installation of PEP**

PEP is typically shipped on one 3.5 inch 1.44 Megabyte floppy disk. To install PEP:

1. Insert the PEP disk into the disk drive.
2. Drag the file PEP.sea to the hard drive that PEP is to be installed on. When the PEP.sea file has been copied to the hard drive you can eject the disk by dragging the "PEP" disk icon to the trash.
3. Double click on the icon of the "PEP sea" file. This will start the installation process by first creating a new folder called "PEP system" on the hard drive and then uncompacting five HyperCard stacks: (1) "Chemical Property Data Base", (2) "PEP Processor", (3) "PEP Help", (4) "PEP Models", and (5) "PEP Batch" (not necessarily in that order).
4. Drag the "PEP.sea" icon to the trash and remove it from your hard drive using the "empty the trash" command which can be found under the Special pull down menu located at the top of the screen.

The installation process is now complete. Test the installation by double clicking on the "PEP system" folder, then double clicking on the file "Chemical Property Data Base". If installation was successful the opening card of PEP will appear. If you have a problem with the installation please contact Mark Holt at (801)750-3916 or Bill Doucette at (801)750-3178. *Note: PEP can also be sent on two 3.5 inch 800k floppy disks if requested.*



## General Programming Description

The PEP software system is a HyperCard™ based program that runs on Apple Macintosh computers. HyperCard, which is bundled with most Macintosh computers sold, offers graphics, information storage, the means to display information in a variety of formats, the ability to establish links between related information, a high level language (HyperTalk), the ability to extend HyperTalk by writing new commands in a compiled language (i.e. C or Fortran) and a mechanism to transfer control to other Macintosh applications. The PEP system uses all these features.

HyperCard treats each screen full of information as a card and each set of related cards as a stack. Cards can contain fields for data and buttons for action procedures to operate on the data in the fields. This allows the standard Macintosh interface to be used without the direct use of the Macintosh toolbox routines, greatly simplifying programming. In order to create a user interface, the programmer simply draws, or creates the buttons or fields that are to be used. The link between buttons, fields and cards is done through HyperTalk. HyperTalk is an high-level, interpreted language used to establish links between related information and perform simple calculations within HyperCard. However, large repetitive tasks and complicated computations can be very slow if HyperTalk is used. HyperCard also allows the programmer to create extensions of HyperTalk in a lower level language. These extensions, called external functions (XFCN) and external commands (XCMD), greatly increase the speed of repetitive and calculation intensive algorithms over using HyperTalk itself and can also be used to implement custom Macintosh features such as popup menus and custom dialog boxes.

## Starting the PEP Software System

If the installation of PEP (as described on the previous page) was successful, a new folder called "PEP system", containing five HyperCard stacks ("Chemical Property Data Base", "PEP Processor", "PEP Help", "PEP Models", and "PEP Batch"), should appear on your hard drive.

PEP is started from the Macintosh operating system by clicking twice (double clicking) on any one of the five stack icons, except the PEP batch stack, shown in Figure 1. The five stacks must be in the same folder on a hard disk. This will open the PEP Processor stack and display the opening card shown in Figure 2. PEP can also be used by opening any HyperCard stack and then choosing "Open" from the "File" menu and selecting the "PEP Processor" stack.

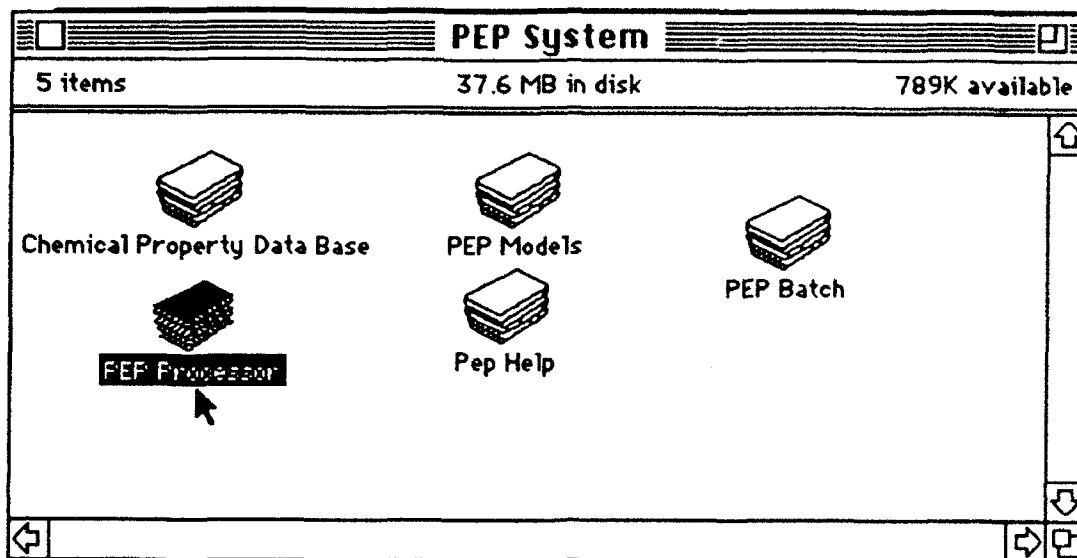


Figure 1. PEP stack icons.

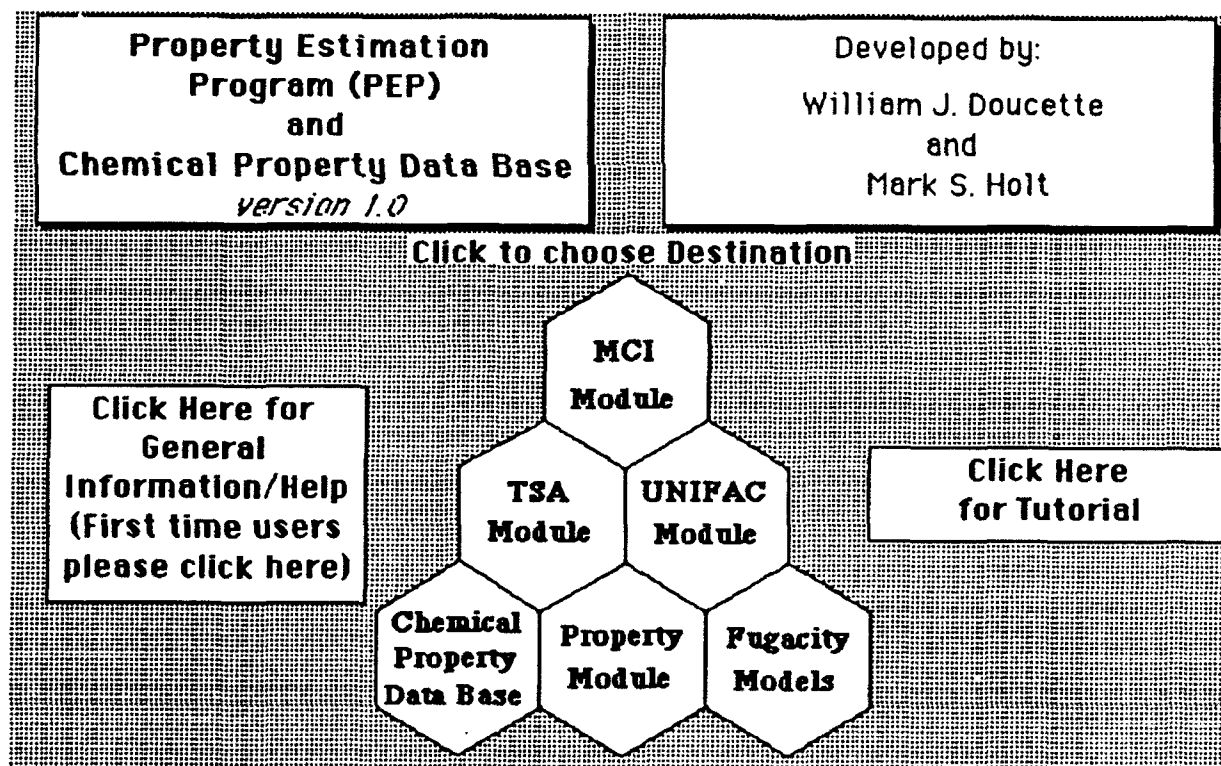
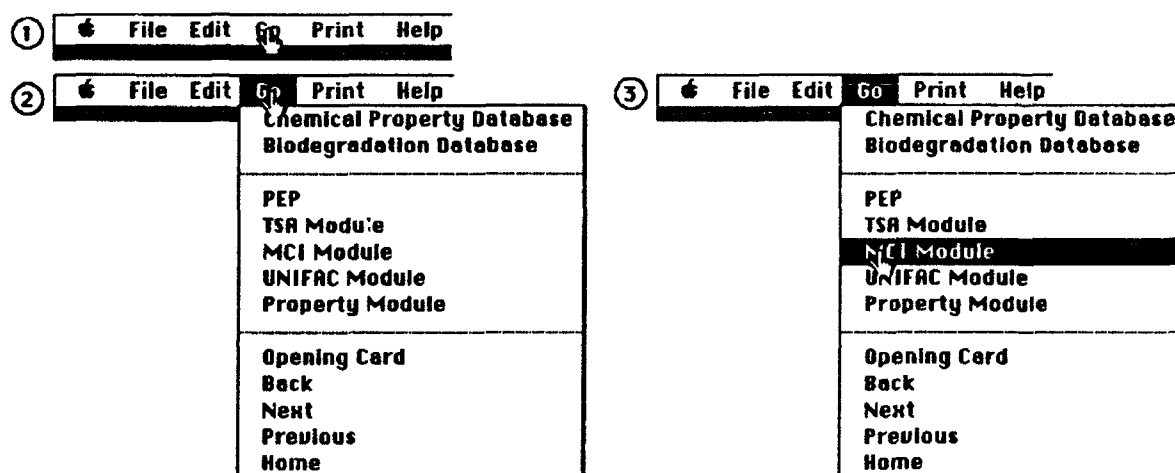


Figure 2. PEP opening screen.

## Menus, Buttons and Icons

In Macintosh applications most cursor movements are accomplished by use of the mouse or trackball. Action buttons are operated by positioning the "hand" cursor over the button area and then depressing and releasing the mouse button. This is referred to as clicking. Popup and pull down type menus are operated by positioning the hand cursor over the button and then depressing and holding the mouse button. While holding down the mouse button move the cursor over the desired menu item and release (Note: on slower machines such as the Mac Plus, the menu selections may be slow to appear, but be patient, they will eventually show up.). Figure 3 shows the steps involved in using a menu for the selection of a command. The action of moving the mouse while holding down the mouse button is called dragging.



1. Menu Bar before the mouse is pressed
2. Menu Bar after the mouse is pressed
3. Menu Bar after the mouse is moved over the desired command but before the mouse button is released.

Figure 3. The steps for using a menu for command selection

The PEP software system is comprised of five HyperCard stacks that are linked together by various menus and buttons. The pull down menus, located at the top of each card and buttons positioned at various card locations allow the user to navigate through PEP.

The File and Edit menus at the top of each PEP card duplicate the general File and Edit menus in Hypercard (Please refer to the Hypercard manual for complete instructions.) The Go menu allows the user to move to either of the four property estimation modules, the chemical property database, the fugacity model or the batch mode.

In addition to the menu items found at the top of each PEP card, buttons, such as those displayed below, can be found on each of the various PEP cards. Figure 4 shows each button icon, its title and its action.










<b>Icon</b>	<b>Title</b>	<b>Action</b>
	PEP Icon	Shows Opening Screen of PEP
	Return Arrow	Takes you back to the card you were at prior to this one
	Help	Shows Help for the current card
	Information	Shows general information
	Eye	Shows the equations or statistics
	Book	Shows the reference
	First Card	Takes you to the opening card of HyperCard
	Pop Up Button	Lets the user choose from a popup menu list
	Action Button	Initiates some action or calculation

Figure 4. Buttons and icons used in PEP.

## Tutorial

The "Click Here for Tutorial" button, accessible on the opening screen of PEP, takes the user to the opening screen of the tutorial. As shown in Figure 5, this screen takes the form of a flow chart depicting the overall design of PEP. Individual tutorials, available for the chemical property database and each of the four property estimation modules, can be activated by clicking on the appropriate button in the flow chart. Each tutorial automatically runs the chosen component of the PEP system while illustrating its operation in a step-by-step manner.

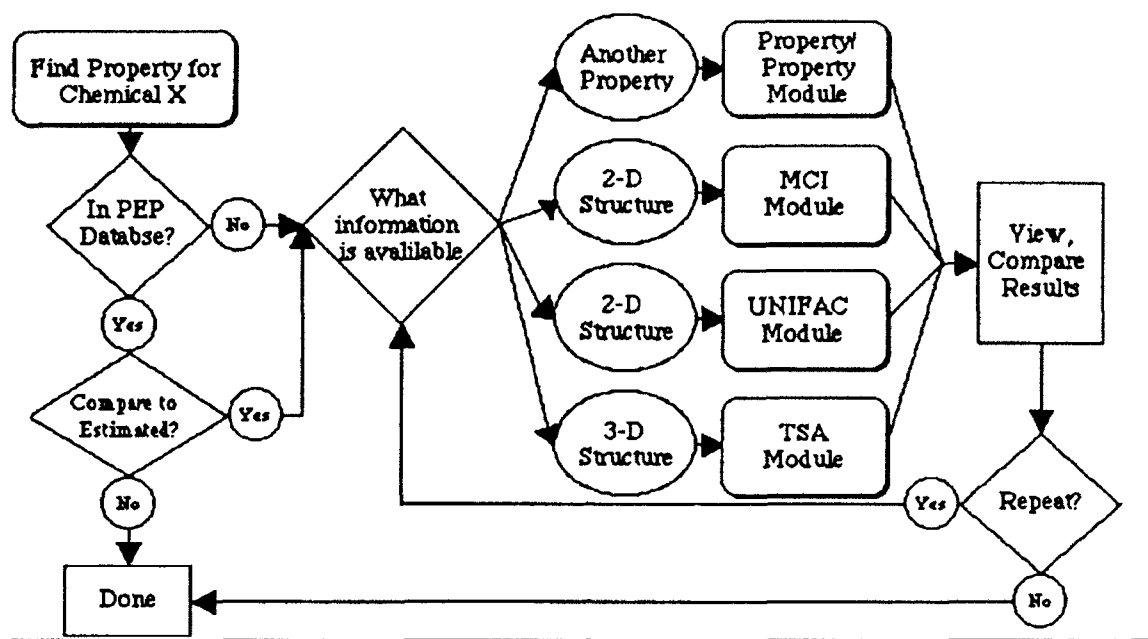


Figure 5. Opening screen of PEP tutorial (Flow chart illustrating overall design of PEP system).

---

## REFERENCE SECTION

### PEP Processor

The PEP Processor stack contains the algorithms for data input, calculations and output of the physical-chemical properties estimated using MCI-Property, TSA-Property, and Property-Property correlations and UNIFAC derived activity coefficients. The PEP Processor is divided into four modules, one module for each of these estimation methods.

The user interface for each module is designed as a flow chart consisting of a series of numbered steps or actions required to estimate the physical-chemical property(s) of interest. The steps are numbered sequentially from left to right and top down. Initially, all but the first step is dimmed. As the user completes the first step the next step will darken, indicating the appropriate progression. However, during any step a darkened, previously accomplished step can be redone or the selection changed. Figure 6 shows this progression of steps.

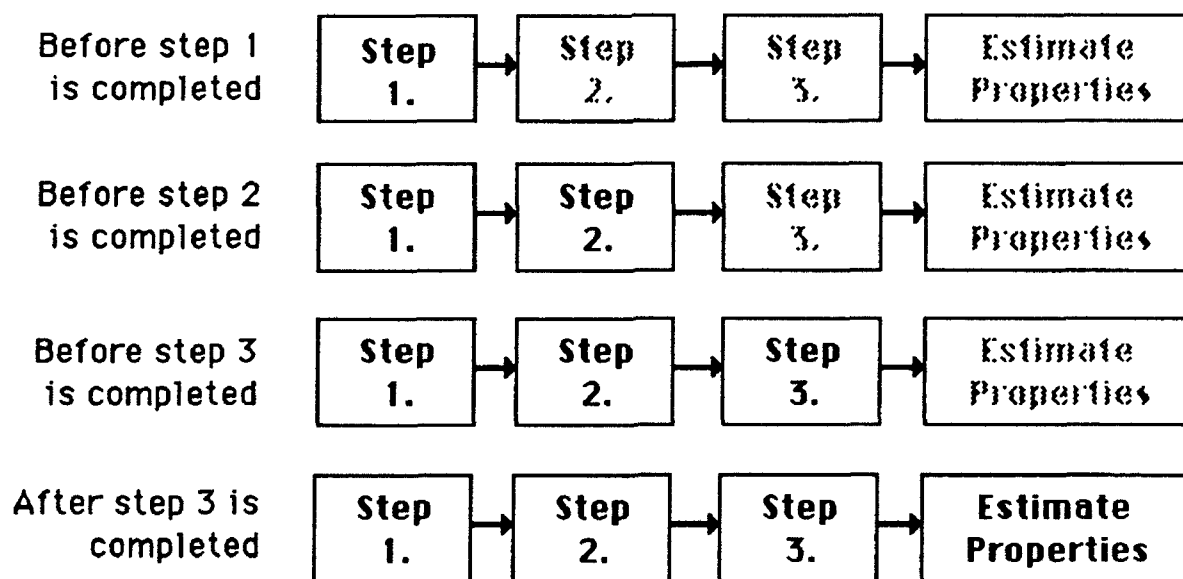


Figure 6. Example illustrating the use of the PEP Processor's flow chart interface.

## MCI Module

### Overview

The user interface of the MCI module, like the other property estimation modules in PEP, is designed in the form of a flow chart depicting the steps that must be completed in order to use the module. As illustrated in Figure 7 by the fully darkened MCI module card, the four steps that must be completed before the selected physical-chemical properties can be estimated using are: (1) input the necessary structural information for the chemical of interest using either SMILES notation or a connection file, (2) calculate and display the MCIs, (3) select the properties to estimate and (4) choose the most appropriate MCI-property regression model.

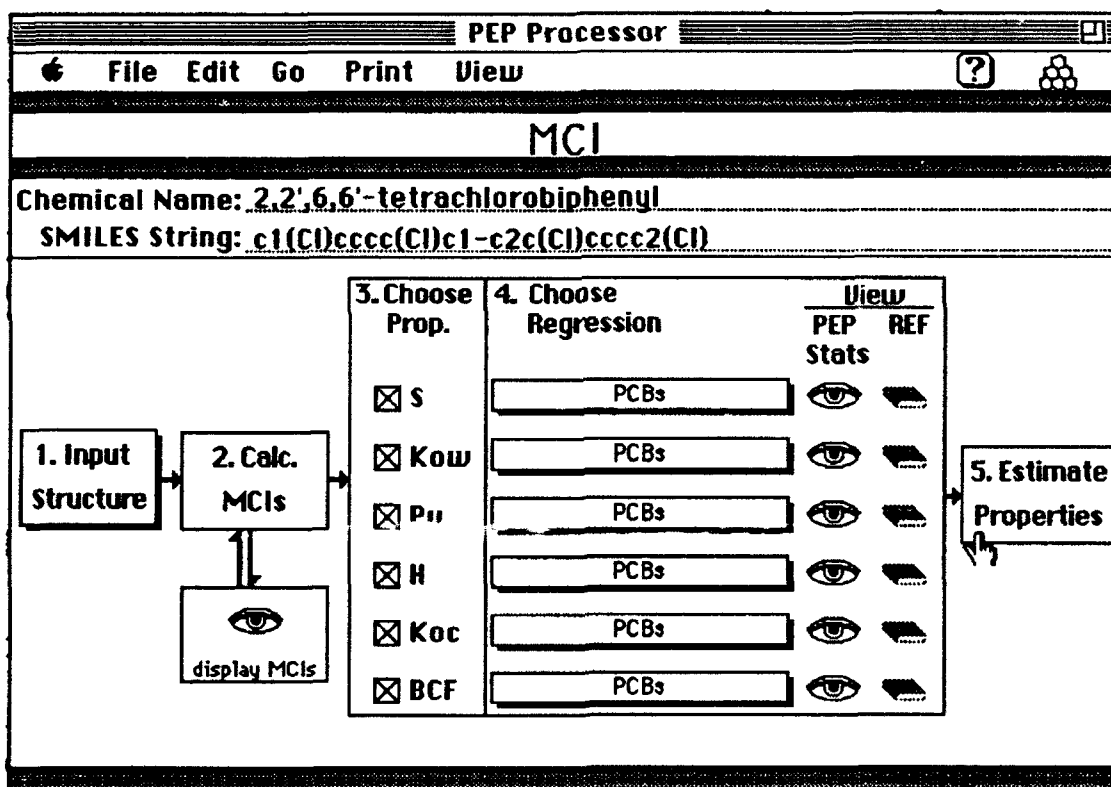


Figure 7. Screen display of PEP MCI module.

## Entering Chemical Structure

Upon entering the MCI module, only the first step in the flow chart is active as indicated by its darkened status. The two dimensional molecular structure of the chemical of interest is needed to calculate the MCIs. You must first input the necessary structural information using either SMILES [33,34] notation or connection files before you can continue to the second step. Select either option from the "Input Structure" popup button.

If you select SMILES, two blank lines appear, one for the chemical name (optional) and one for the SMILES string. Once you have entered the SMILES string, click the "OK" button or the carriage return. The SMILES string must conform to the standard set by Anderson, Veith, and Weininger (1987) and Weininger (1988) with the following exceptions: a single bond connecting aromatic rings must be explicitly denoted by an "-", the SMILES string must be Hydrogen suppressed, and the SMILES string cannot contain any "{}" or "[]" qualifiers. SMILES is a chemical notation language specifically designed for computer use. It is a method of "unfolding" a 2D chemical structure into a single line of characters containing the structural information. For users unfamiliar with SMILES notation, a detailed description describing its use can be found in scrollable window directly below the SMILES string input line.

If a connection table is chosen as the input method, the standard Macintosh file selection dialog box, as shown in Figure 8, will be used to select the file. This requires that a connection table has already been created for the chemical of interest.

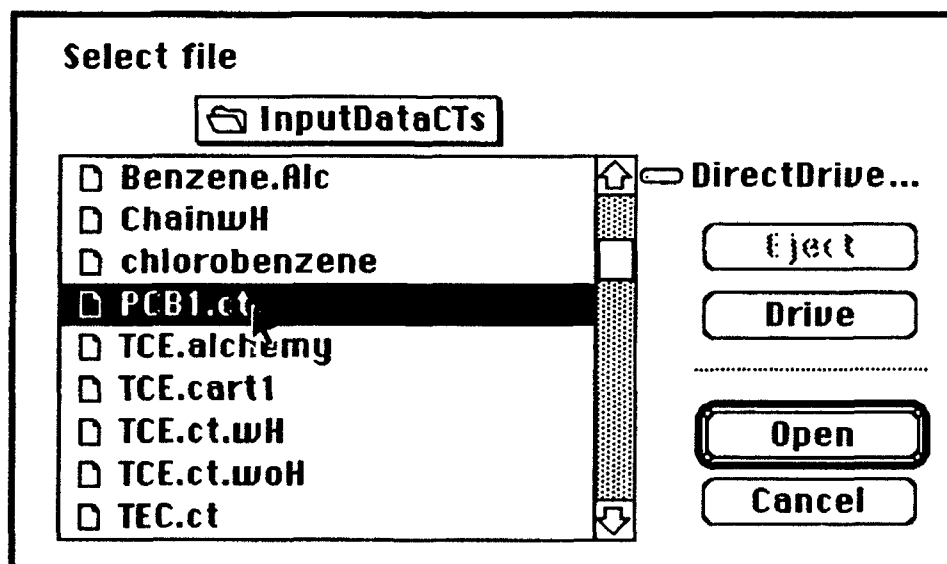


Figure 8. Standard Macintosh file selection dialog box.



Connection tables can be generated from commercially available, two-dimensional (2D) chemical drawing programs, such as ChemDraw™, Chemintosh™, or ISIS/Draw™, that have the ability to save the structure as a connection table file. The connection table file must be formatted the same as a connection table from ChemDraw (1989). An example of a ChemDraw compatible connection table is shown in Figure 9.

Title line	Example connection table			
Number of Atoms	10	10		
Number of Connections	7.62500	2.29167	0.00000	C
	7.62500	3.12500	0.00000	C
X,Y,Z Coordinates (Not used)	8.33333	3.54167	0.00000	C
	9.04167	3.12500	0.00000	C
	9.04167	2.29167	0.00000	C
Atom Symbol	8.33333	1.87500	0.00000	C
	8.33333	1.04167	0.00000	C
	9.16667	1.04167	0.00000	C
	10.00000	1.04167	0.00000	C
	10.83333	1.04167	0.00000	C
	2	1	4	1
	3	2	4	1
	4	3	4	1
Atom Numbers	5	4	4	1
	6	5	4	1
Bond Type	1	6	4	1
Not Used	7	6	1	1
	8	7	1	1
	9	8	2	1
	10	9	1	1

Figure 9. Example connection table.

The first data line contains the title of the chemical or any other identifier. The second line consists of two numbers separated by a comma. The first number is the number of atoms in the connection table and the second is the number of connections described in the connection table. The remaining lines describe the type of atoms in the molecule and their location (atom information) and the atoms to which they are connected (connection information). The total number of lines depends on the number of atoms in the molecule and the number of connections.

Atom information is contained in four columns separated by one or more spaces. Columns one through three are the X,Y,Z coordinates of the atom (not used for calculation of MCIs) and column four contains the atom symbol (e.g. C, Cl, Br, N etc.). Connection information is contained in four columns of whole numbers. Columns one and two contain the atom numbers of the atoms that are connected (atoms are numbered consecutively), column three contains the bond type (1, 2, 3, or 4), and column four usually contains a 1 (not used). The bond type in column three can be either "1" for a single bond, "2" for a double bond, "3" for a triple bond, or "4" for an aromatic bond. (NOTE: *The Macintosh compatible molecular modeling program, Alchemy II™, generates files containing connection table information along with 3D atomic coordinates. To use alchemy files for input into the MCI module simply treat the alchemy file as a "ChemDraw Connection Table".*)

As described in the next section, MCIs are calculated from the hydrogen suppressed structure of a chemical. Consequently, for the most efficient calculation of MCI, the connection tables created for input into the MCI module should be created from hydrogen suppressed structures. If the connection tables are not hydrogen suppressed PEP will automatically remove them. This can result in a significant increase in the time required to calculate the MCIs.

### Calculating MCIs

As soon as the structure is entered, the second step becomes active. Clicking on the "Calc. MCIs" button starts the calculation of the MCIs. If the structure was entered as a SMILES string the "mciSmile" XFCN converts the SMILES string into the proper format for the "mcichi" XFCN which calculates the MCIs. Similarly, if a connection table was used for input the "mciConvert" XFCN converts the connection table into the proper format. The MCIs that are calculated are listed along with the variable names in Table 1. Once the MCIs have been calculated, they can be viewed, exported, or printed by using the "Display MCIs" button under the "Calc. MCIs" button.

Table 1. MCIs calculated by PEP.

Variable Name	MCI Title	Orders calculated
np	normal path	0 through 6
ncl	normal cluster	3 through 6
nch	normal chain	3 through 6
npc	normal path/cluster	4 through 6
bp	bond path	0 through 6
bcl	bond cluster	3 through 6
bch	bond chain	3 through 6
bpc	bond path/cluster	4 through 6
vp	valence path	0 through 6
vcl	valence cluster	3 through 6
vch	valence chain	3 through 6
vpc	valence path/cluster	4 through 6
$\Delta vp$	delta valence path	0 through 6

### How PEP calculates MCIs

To calculate the MCIs for a given compound, a delta (d) value are first assigned to each non-hydrogen atom in the structure. Three d values were computed in this study: normal, bond, and valence. Normal deltas are computed by summing the number of bonds (single, double, etc. are counted as one bond) connected to the atom whose delta is being calculated. The bond deltas are calculated the same as the normal deltas except the bonds were taken at their face value (single is one, double is two, etc.) instead of each bond being equal to one. Valence deltas for each atom are computed according to equations (1) and (2) (Kier and Hall, 1986):

$$d_v = Z_v - h \quad (1)$$

$$d_v = (Z_v - h)/(z - Z_v) \quad (2)$$

where  $d_v$  is the valence delta,  $Z_v$  is the number of valence electrons in the atom,  $h$  is the number of hydrogen atoms bound to the atom, and  $Z$  is the atomic number of the atom. Equation (1) is used for those atoms in the first row of the periodic chart, and equation (2) is used for all other atoms.

Once the delta values have been calculated for each atom in the molecule, simple, bond and valence indices of different orders and types can be calculated. The order refers to the number of bonds in the skeletal substructure of fragment used in computing the index: zero order defines individual atoms, first order used individual bond lengths, second order uses two adjacent bond combinations, and so on. The type refers to the structural fragment (path, cluster, path/cluster or

chain) used in computing the index. A more detailed explanation of the calculation of MCIs can be found in Kier and Hall (1986).

The MCI calculation routine in PEP calculates simple, bond and valence indices of several types (path, cluster, chain, and path/cluster) and orders (0 through 6), if possible, for each molecule, resulting in a maximum of 54 index values for each molecule.

To account for non-dispersive force effects on aqueous solubility and solubility related properties zero through six order  $\Delta$  valence path indices ( $\Delta\chi$ ), as described by Bahnick and Doucette (1988), are calculated by PEP, in addition to the 54 indices described above. To calculate  $\Delta\chi$  indices, a nonpolar equivalent is made by substituting C for O or N atoms. MCIs are calculated for the nonpolar equivalent and values for  $\Delta\chi$  can be computed for each type of index by:

$$\Delta\chi = (\chi)_{np} - \chi \quad (3)$$

where  $\Delta\chi$  is the delta index,  $(\chi)_{np}$  is the index for the non-polar molecule and  $\chi$  is the index for the original molecule.

### Choosing the Properties

After the MCIs have been calculated, the third step becomes active. Select the property or properties you would like to estimate by clicking on the check box button next to it. You can simultaneously select all the properties by holding down the shift key and clicking any one of the property buttons.

### Choosing the Regression Models and Chemical Classes

After the properties are selected, step four becomes active and the regression models available for each property are displayed in a popup menu. Two categories of MCI-property relationships are displayed for each property. The first category of MCIs property relationships, preceded with the word PEP, were developed in this project using the experimental values reported in the PEP property database. "Universal" MCI-property relationships were developed using all available experimental data for a given property regardless of chemical class. "Class-specific" MCI-property relationships were developed if property values were available for a sufficient number (10 or greater) of compounds within a particular chemical class (i.e. PCBs, PAHs, ureas). In addition, several multi-class MCI-property correlations were developed for more broad classes of compounds such as: halogenated aliphatics and halogenated aromatics.

The second category of MCI-property relationships displayed for each property were obtained directly from the literature and are located below the PEP relationships, separated by a gray line, in the popup menu. By clicking the "book" found at the left of each literature MCI-property regression model, the coefficients  $r^2$  value and the appropriate citation can be displayed

To illustrate the potential hierarchy of MCI-property relationships available to the user, an example for the predicting the vapor pressure ( $P_v$ ) of a polychlorinated biphenyl (PCB) is provided below. There are three appropriate PEP-derived MCI-property relationships available to the user, one developed using only PCBs, one using halogenated aromatics including PCBs and one using all compound types:

$$\log P_v = 5.814 (\text{nc5}) - 2.428 (\text{np3}) + 9.479 \quad (\text{PCBs})$$

$$\log P_v = -1.559 (\text{bp1}) + 6.622 \quad (\text{Halogenated aromatics})$$

$$\log P_v = -1.275 (\text{np3}) + 5.261 \quad (\text{Universal})$$

Generally, the use of a "class-specific" relationship, if available, should provide the best estimate (i.e. the estimate associated with the least amount of uncertainty). To automatically aid you in choosing the most appropriate MCI-property relationships, PEP looks in the SMILES string or connection file for groups of atoms and bonds that distinguish various chemical classes. The number of MCI-property relationships or chemical classes that are chosen by the program is determined by the number of different distinguishing subgroups that are found. For the example shown above, the most appropriate regression model, PCBs, would be made the default equation. The two other appropriate models, halogenated aromatics and Universal, would be denoted with a ♦ in the popup menu. If the compound entered into PEP does not fit one of the class specific models the "Universal" equation is selected as the default. You may also choose to ignore the regression model chosen by PEP and select your own.

MCI-property regression models are available for the following "classes" of chemicals: Universal, Universal Nonionizable, Universal Ionizable, Alcohols, Arilines, Carbamates, Halogenated, Aliphatics, Nonhalogenated Aliphatics, Halogenated Aromatics, PCBs, PAHs, Phenols, Triazines, and Ureas. Examples of representative chemicals for each of the classes can be found in the View menu. To see the structures of the chemicals click on the class name. Not all of the classes listed above are implemented for each property. The models not available for the properties to be estimated are dimmed in the popup menu.

## Statistics Cards

A summary of the regression statistics and list of compounds used to develop and evaluate each MCI-property relationship can be displayed by clicking the "eye" or "view statistics option" found at the left of each regression model. Information displayed on the statistics card includes: the MCI-property regression equation, the list of chemicals used in developing the regression model, the standard errors of the coefficients in the regression equation, the Analysis of Variance (ANOVA) table, the  $r^2$  value, a graph of the predicted vs. estimated values, a graph of the residuals vs. the predicted values, a graph of the residuals vs. the number of standard deviations, a normal probability plot of the residuals, the  $X'X$  inverse matrix and appropriate reference. An example of the statistical information provided for each MCI-property relationship is shown in Figure 10.

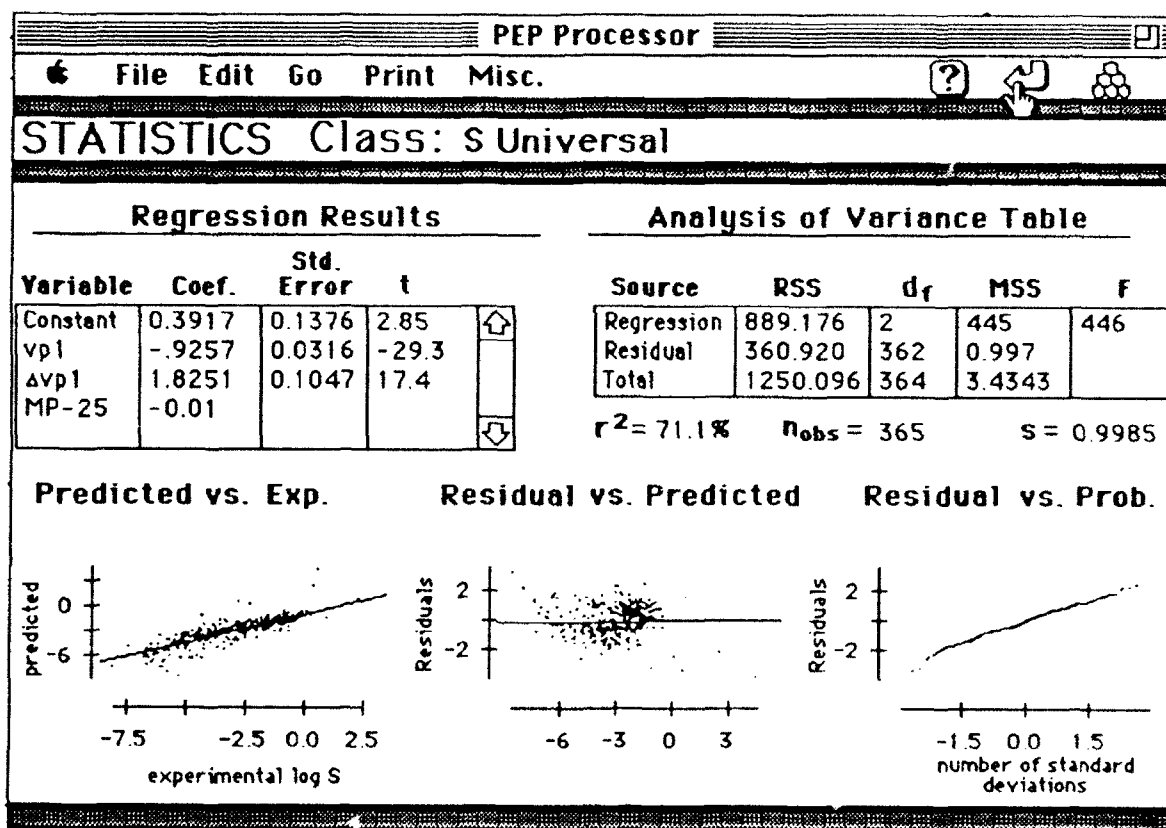


Figure 10. Statistics card associated with the MCI module.

The ANOVA table contains the degrees of freedom, the residual sum of squares, the residual mean square, and the variance ratio (F) for regression, residual and the total source of errors.

The  $X'X$  inverse matrix is used in PEP to calculate the prediction interval of an estimate. The matrix is derived by pre-multiplying the  $X$  matrix by its transpose and then inverting the result. The  $X$  matrix has a column for each variable in the regression equation and a row for each observation used to calculate the regression equation. Each row contains the value used for each of the variables in the regression equation. For example, if the regression equation is

$$Y_j = b_0 + b_1X_{j1} + b_2X_{j2}$$

where  $j$  is 1 to the number of observations,  $Y_j$  is the estimated values for each observation,  $b_i$  are the regression coefficients, and the  $X_{ji}$  are the values of the variables used then the  $X$  matrix is

1	$X_{11}$	$X_{12}$
1	$X_{21}$	$x_{22}$
:	:	:
:	:	:
1	$X_{j1}$	$X_{j2}$

The resulting  $X'X$  inverse is a square matrix with the number of rows and columns equal to the number of variables in the regression equations.

If the QSPR or QPPR was taken from the literature only the input variables and the statistical information provided in the original reference is included.

### Estimating the Properties and Viewing the Results

Click the "Est. Property" button to calculate the selected properties. The results card displays the estimated properties and their respective 95% prediction interval *Note: 95% prediction intervals are not available for MCI-property relationships taking from the literature.* You can return to the previous card by clicking on the "return" button at the upper right corner of the results card. The "Go" menu can be used to move to another module. You can compare the estimate property values with those contained in the "Chemical Property Database", if available, by clicking on the "Look in DB" button. Clicking this button activates a database "search by name" routine. The name on the results card must match exactly the name in the database for the search routine to find the compound. If a property value is not found an NA will be displayed.

### Adding or Deleting MCI-Property Regression Models

Additional MCI-property regression models can be added to the PEP MCI module through the statistics cards. Choosing the "New Stat. Card" option from the "Misc." menu of a statistics cards

while the type of statistics card that is to be added is the current card The statistics for a new regression equation can be added to PEP b. First the new card must be titled. This title will be used in the popup menus and on the results card. The first 28 characters of the title must be unique. The second step is to enter the regression equation. The user will be prompted for the number of terms in the equation, then prompted for each coefficient and the associated variable with the dialog box shown in Figure 11. The variables which are available for that type of statistics card will be in a popup menu for easy, consistent selection.

**Choose MCI and Input Coefficient**

Choose MCI :

Coefficient:

Figure 11. Example dialog box for the input of new MCI-property relationships

A relationship can easily be deleted from PEP by first making that statistic card the current card, and then choosing "Del. Stat. Card" from the "Misc." menu. The user will be prompted to confirm the deletion and then the regression list will be rebuilt.

### Limitations of MCI-Property Regression Models

Selection of the most appropriate MCI-property relationship depends on the structure of a particular chemical of interest, knowledge of the mechanism of the process, and the extent of the database used to develop the MCI-property relationship. For example, some MCI-property relationships are broader than others in the range of chemicals that are covered, and some have been established with a better understanding of the mechanisms or properties involved.

One problem that has limited the widespread acceptance of MCI-property correlations is that the actual physical meaning associated with the individual indices is not well understood. Frazier



(1990) and Doucette and Holt (1991), however, have shown a strong correlation between calculated molecular surface area and several MCIs for a variety of organic chemicals. MCI-property correlations tend to be class specific and thus are highly dependent on the type and range of compounds that were used to derive a particular correlation. Indiscriminate use of such models without an examination of number and type of compounds used to develop the model can result in considerable error.

## TSA Module

### Overview

As shown in Figure 12, the TSA module is similar in design to the MCI module. However, unlike molecular connectivity, the calculation of TSA requires information describing the geometry of the molecule in terms of its 3-D atomic coordinates.

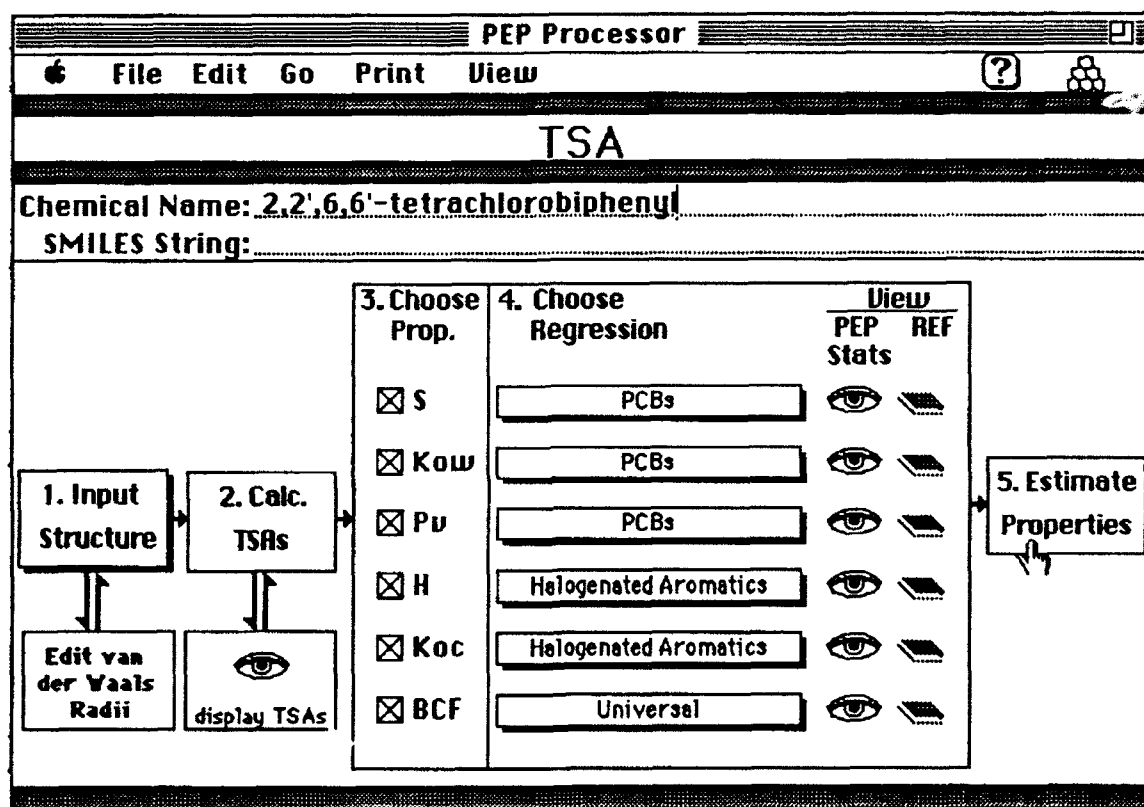


Figure 12. TSA module card from PEP

Three-dimensional coordinates can be obtained from X-ray crystallography data or from molecular modeling. Alchemy from Tripos Associates (1989) and Chem3D+ from Cambridge Scientific (1989) are examples of Macintosh compatible molecular modeling software that allows the user to draw a chemical structure, energy minimize the structure, and produce a file containing the three-dimensional coordinates. The TSA module is also designed to accept files generated by other hardware/software combinations including UNIX or VAX versions of CONCORD (Tripos

Associates, Inc., 1990), a hybrid expert system and molecular modeling software designed for the rapid generation of high quality approximate 3-D molecular structures.

### **Entering the Structural Information**

To estimate properties using the TSA module, you must first input the necessary structural information by using one of the three options available under the popup menu titled "Input Structure": (1) Alchemy file, (2) cartesian coordinates file, or (3) manually entered cartesian coordinates. The preferred method of structural input is via Alchemy files because they contain both the three-dimensional structure and the connection information. This information allows PEP to calculate the chemical's TSA and determine the most appropriate TSA-property relationship based on chemical class. Therefore, if an Alchemy file is chosen for the input then only the file selection dialog box will be presented for selection of the file. However, if either of the cartesian coordinate options are selected for input, both the standard file selection dialog box will be presented and an opportunity to select the connection table file or enter a SMILES string will be presented.

The format of an Alchemy file, shown in Figure 13, is similar to that of a ChemDraw connection table discussed previously. The first line of an Alchemy file contains the number of atoms followed by "ATOMS," the number of bonds followed by "BONDS," the number of charges followed by "CHARGES," and then the title of the file. The next set of lines contains six columns of information for each atom in the molecule including the hydrogen atoms. Column one contains the atom numbers, column two contains the atomic symbols, and columns three through five contain the X,Y,Z coordinates of the atom. The set of lines describing the atoms is followed by a series of lines containing four columns describing the bonds. The first column contains the bond number, the second and third columns contain the atom numbers of the two atoms connected, and the fourth column contains the type of bond either "SINGLE," "DOUBLE," "TRIPLE," or "AROMATIC."

	Number of Atoms	Number of Bonds	not used		title
	6 ATOMS, 5 BONDS, 0 CHARGES, TCE				
Atom numbers	1	C2	0.0008	0.0149	0.0048 0.0000
Atom symbols	2	CL	0.7042	1.5686	0.4047 0.0000
X,Y,Z coordinates	3	C2	0.6631	1.1153	0.2713 0.0000
	4	CL	1.5811	0.0700	0.7463 0.0000
not used	5	CL	2.2442	1.0786	1.0242 0.0000
	6	H	0.2306	2.0846	0.0245 0.0000
Bond numbers	1	1	2	SINGLE	
	2	1	3	DOUBLE	
Atom numbers	3	1	4	SINGLE	
Bond type	4	3	5	SINGLE	
	5	3	6	SINGLE	

Figure 13. Example Alchemy file.

PEP accepts cartesian coordinates files having the following format. The file has one header line indicating the number of atoms in the molecule. The rest of the lines in the file describe each atom in the molecule. Only the first five columns of each line are used. Column one contains the atom symbols, column two contains the atom numbers, and columns two through five contain the X,Y,Z coordinates of the atom. An example cartesian coordinates file is shown in Figure 14.

Number of Atoms	Atom Symbols	Atom Numbers	X Y Z coordinates			Not used	
6							
	Cl	1	1.658981	1.412048	0.220383	12	2
	C	2	0.678864	0.010635	0.004303	2	1
	C	3	0.661346	0.015198	0.109131	2	2
	Cl	4	1.498978	1.476791	0.333939	12	3
	Cl	5	1.607437	1.456451	0.031769	12	3
	H	6	1.277542	0.912277	0.055084	5	2

Figure 14. Example Cartesian coordinate file.

If the option to manually enter the coordinates is chosen then a card is presented, as shown in Figure 14, that allows the coordinates and atom symbol of each atom to be entered individually from the keyboard. The X,Y,Z, coordinate values and atom symbols for each atom in the molecule are entered in the appropriate labeled boxes one at a time. After the information for each line is correctly entered the "Line OK" button is clicked. This enters the information into the scrollable window below. This process is repeated for each atom in the molecule. When all the atoms have been entered clicking the "Done" button will send the structural information to the TSA module.

**File Edit Go Print Misc.** ?

---

**Enter XYZ coordinates for the chemical**

---

**Chemical Name:** Benzene

**Fill in one atom at a time at the bottom, type Return or Tab to put it into the table. To Edit an entry click on the line in the table. Click "All Done" when the table is complete.**

Atom	Symbol	X coord	Y coord	Z coord
2	C			

Line OK

1 C 1 1 1

Clear  
Cancel  
All Done

Figure 15. Example card for the entry of atomic coordinates.

### Calculating Total Surface Area (TSA)

In addition to the 3-D molecular structure, the user must also input van der Waals radii for each of the atoms before the TSA of the molecule can be calculated. PEP automatically enters a van der Waals radius for each atom using values from Pauling (1960). However, when the "Calc. TSA" button is clicked the user has the opportunity to edit the van der Waals radii using the dialog box shown in Figure 16.

Atom Symbol	Radius Å
C	1.7
CL	1.8
BR	1.95
O	1.4
N	1.5
H	1.2
S	1.85
P	1.9
F	1.35

Default values taken from:  
Pauling, 1960. "The Nature  
of the Chemical Bond."  
Cornell University Press.  
Ithaca, New York.

**Reset Defaults**

**Cancel**

**OK**

**Solvent  
Radius**  Enter 0.0 for Total Surface  
Area to be calculated

Figure 16. Dialog box for editing van der Waal radii.

This dialog box also contains a place to enter a solvent radius. If it is left at "0.0" then the total surface area will be calculated. Some relationships from the literature require the solvent accessible surface area to be calculated. If this is the case, the desired solvent radius can be entered.

Once the molecular geometry and the van der Waal radii are input, the "Calc. TSA" button becomes active and TSA can be calculated using a XFCN that was adapted from the SALVO2 algorithm developed by Pearlman (1980). This algorithm represents each atom of a molecule by a sphere centered at the equilibrium position of the nucleus. The radius of the sphere is equal to that of the van der Waals radius. Planes of intersection between spheres are used to estimate the contribution to surface area from the individual atoms or groups. The program computes the surface area of individual atoms or group by numerical integration, and the overlap due to intersecting spheres is excluded from the calculation. TSA is calculated by the summation of individual group contributions. These areas are then imported to the "TSAs" card shown in Figure 17.

File Edit Go Print Misc. ? ↩

---

## TSAs

---

**Chemical Name:** Benzene.Alc

TSAs Calculated with 0.0 Solvent radius						Using a 0.0 Solvent Radius
atom #	sym.	isolated area	accessible area	isolated volume	accessible volume	
1	ArC	36.317	15.391	20.580	12.027	total area (Å <sup>2</sup> ) 92.348
2	ArC	36.317	15.397	20.580	12.028	total volume (Å <sup>3</sup> ) 72.158
3	ArC	36.317	15.382	20.580	12.023	
4	ArC	36.317	15.394	20.580	12.027	
5	ArC	36.317	15.390	20.580	12.027	
6	ArC	36.317	15.394	20.580	12.026	

N-TSA	O-TSA	P-TSA	S-TSA	ArN-TSA
0	0	0	0	0

**Not Using a 0.0 Solvent Radius**

total area (Å<sup>2</sup>)  
Not Calc.

total volume (Å<sup>3</sup>)  
Not Calc.

Figure 17. Example TSA card.

To quantify the surface area attributed to the polar portions of the molecule the surface areas of nitrogen, oxygen, phosphorous, sulfur, and aromatic nitrogen atoms are individually separated from the TSA and placed on the "TSAs" card. A more detailed description of the TSA calculation method is provided by Pearlman (1980).

### Choosing the Most Appropriate TSA-Property Regression Model

After the TSA has been calculated, you can display the values and/or choose the properties of interest and a corresponding regression model using the same approach described in the MCI module. As discussed previously, "class-specific" regression models generally yield estimates associated with the least amount of uncertainty. If an Alchemy file is used to enter the 3D structure information or if a SMILES string or connection file is entered along with the cartesian coordinates, the decision support system in PEP will choose the most appropriate TSA-property regression model(s) as described in the MCI module. The most appropriate regression model will be made the default and a ♦ will be placed next to the other appropriate class name in the popup menu containing a list of the regressions. If no class specific regression models are available, the

"Universal" equation will be made the default. The operation of the TSA module from this point on is identical to that of the MCI module. Note: for solid solutes the melting point is needed to estimate the solubility. When the "S" button is clicked, the user is prompted with a dialog box to enter the melting point. Because the solubility will be estimated at 25° C, only if the melting point is above 25° C (solid solute) will the value be required. If the melting point is below 25° C it is set equal to 25° C. At this time only the " $\leq 25^\circ \text{C}$ " and "known" options are useable in the dialog box.

TSA-property regression models available within PEP are: Universal, Universal Nonionizable, Universal Ionizable, Alcohols, Anilines, Carbamates halogenated Aliphatics, Nonhalogenated Aliphatic, Halogenate Aromatics, Nonhalogenated Aliphatic, Halogenate Aromatics, PCBs, PAHs, Phenols, Triazine, and Ureas. Not all models are available for each property. The models not available for the properties to be estimated are dimmed in the popup menu. You can also view the statistical information associated with each model by clicking on the "eye" next to the regression equation.

### **Estimating the Properties**

After the appropriate TSA-property relationships have been selected you can now click on the "Estimate Property" button to calculate the estimated properties. The estimated properties and their respective 95% prediction interval will be displayed on the results card. Return from the results card by clicking on the "Return arrow" button at the upper right hand corner of the results card. Use the "Go" menu to move to another module. View values in the Chemical Property Database by clicking on the "Look in DB" button. If a property value is not available in the database, an NA will be displayed in the property field.

### **Development of TSA-Property Relationships**

The PEP TSA-Property relationships that were were developed using stepwise regression techniques and the data in the Chemical Physical Data Base . The stepwise regression was stopped when the coefficient of determination did not improve by at least 0.05 when the next variable was added.

For compounds containing polar functional groups, the addition partial TSA terms (i.e. nitrogen (N-TSA), oxygen (O-TSA), aromatic nitrogen (ArN-TSA), sulfur (S-TSA), or phosphorous (P-TSA)) significantly improved the TSA-property regression models. To view the regression information by clicking on the "eye" next to the regression title on the results card or on the TSA card.



## **Limitations of TSA-Property Relationships**

A major factor in the solubiliztion process is the energy required to create a cavity in the solvent into which the solute is placed. The energy needed for the hole formation is considered to be proportional to the surface area of the solute. TSA has been found to be linearly related to the logarithm of solubility for many classes of non-ionizable organic chemicals.

As with the MCI module, selection of the most appropriate TSA-property relationship depends on the structure of a particular chemical of interest, knowledge of the mechanism of the process, and the extent of the database used to develop the TSA-property relationship. Some TSA-property relationships are broader than others in the range of chemicals that are covered and some have been established with a better understanding of the mechanisms or properties involved.

## UNIFAC Module

### Overview

The UNIFAC (UNIQUAC Functional Group Activity Coefficient) group contribution method for calculating activity coefficients, as described by Grain (1990), is implemented using both HyperTalk and an XFCN. The functional group interaction parameters, presented by Gmehling et al. (1982) and derived from vapor-liquid equilibria (VLE), are used in the calculation routine but can be changed by the user. After the activity coefficients are calculated they can be displayed along with relevant intermediate values and used to estimate S and  $K_{OW}$  by the following expressions (Arbuckle, 1986):

$$K_{OW} = 0.115 \gamma^{\infty w} / \gamma^{\infty o} \quad (2)$$

$$S \text{ (mol/L)} = 55.6 / \gamma^{\infty w} \quad (3)$$

where  $\gamma^{\infty w}$  is the activity coefficient of the chemical infinitely dilute in water and  $\gamma^{\infty o}$  is the activity coefficient of the chemical infinitely dilute in octanol.

The operation of the UNIFAC module, illustrated in Figure 18, is considerably different than the correlation modules previously described. To use the UNIFAC module the you must input structural information, calculate the activity coefficients, choose then choose the properties to be estimated. Currently the only properties that can be directly estimated using the UNIFAC module are S and  $K_{OW}$ .

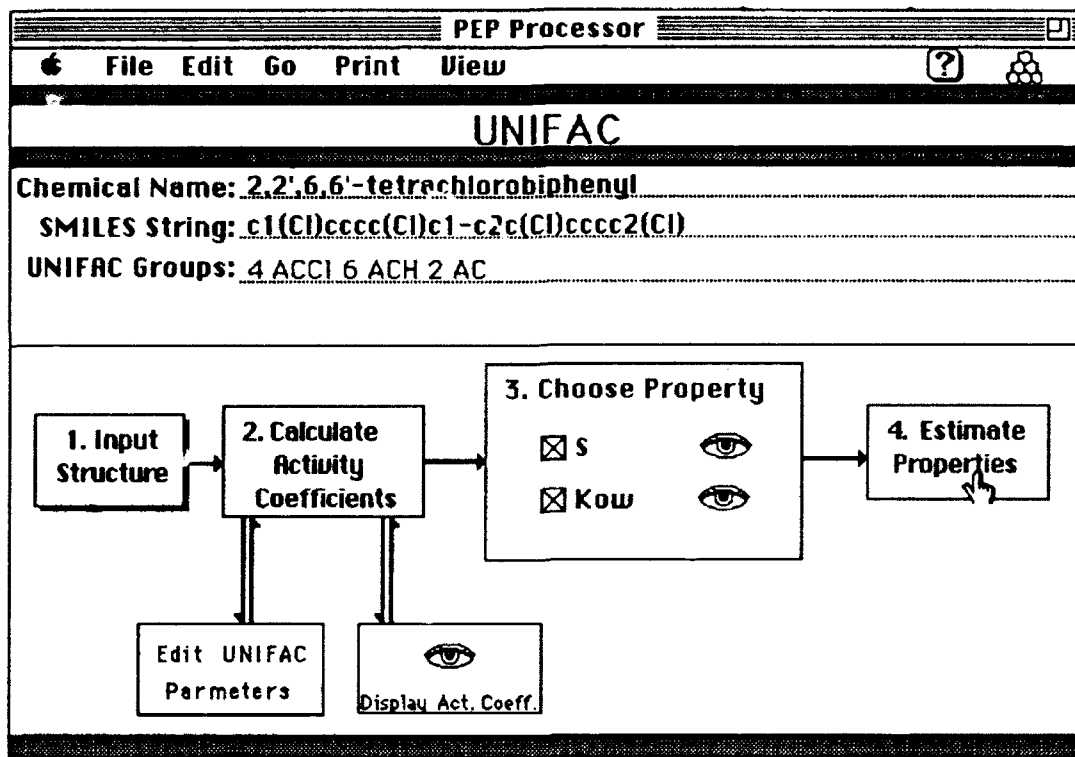


Figure 18. Example card from the PEP UNIFAC module.

### Entering Structural Information

Calculation of activity coefficients via the UNIFAC approach requires that the user input the valid UNIFAC groups that make up the chemical of interest and the number of each group present. PEP provides the user with three options for entering the appropriate UNIFAC groups using the popup menu under the button titled "Input Structure": (1) hand selecting the groups from a list in HyperCard, (2) using a connection table, and (3) using SMILES.

If the connection table option is chosen, a the standard file selection dialog box is presented where the user can select the desired connection table file. The "Struct" XFCN then uses the connection information to dissect the molecule into its UNIFAC groups.

If SMILES is chosen as the input method, the user is then prompted to input the chemical name and the SMILES string. The Struct XFCN uses the connection information contained in the SMILES string to build the UNIFAC input string.

When the user chooses to hand select the UNIFAC groups, a card is displayed showing the first 37 groups as shown in Figure 19.

### Main Groups

(1) $\text{CH}_2^-$	(4) $\text{ACCH}_2^-$	(10) $\text{CHO}$	(15) $\text{CNH}$	
<input type="radio"/> $\text{CH}_3$	<input type="radio"/> $\text{ACCH}_3$	(11) $\text{CCOO}^-$	<input type="radio"/> $\text{CH}_3\text{NH}$	
<input type="radio"/> $\text{CH}_2$	<input type="radio"/> $\text{ACCH}_2$	<input type="radio"/> $\text{CH}_3\text{COO}$	<input type="radio"/> $\text{CHNH}$	
<input type="radio"/> $\text{CH}$	<input type="radio"/> $\text{ACCH}$	<input type="radio"/> $\text{CH}_2\text{COO}$	<input type="radio"/> $\text{CH}_2\text{NH}$	
<input type="radio"/> $\text{C}$				
(2) $\text{C}=\text{C}^-$	(5) $\text{OH}$	(12) $\text{HCOO}$		
<input type="radio"/> $\text{CH}_2=\text{CH}$	(6) $\text{CH}_3\text{OH}$	(13) $\text{CH}_2\text{O}^-$	(16) $\text{(C)}_3\text{N}^-$	<input type="button" value="More"/>
<input type="radio"/> $\text{CH}=\text{CH}$	(7) $\text{H}_2\text{O}$	<input type="radio"/> $\text{CH}_3\text{O}$	<input type="radio"/> $\text{CH}_2\text{N}$	
<input type="radio"/> $\text{CH}_2=\text{C}$	(8) $\text{ACOH}$	<input type="radio"/> $\text{CH}_2\text{O}$	<input type="radio"/> $\text{CH}_3\text{N}$	<input type="button" value="DONE"/>
<input type="radio"/> $\text{CH}=\text{C}$		<input type="radio"/> $\text{CH-O}$		
<input type="radio"/> $\text{C}=\text{C}$	(9) $\text{CH}_2\text{CO}^-$	<input type="radio"/> $\text{FCH}_2\text{O}$	(17) $\text{ACNH}_2$	<input data-bbox="1377 934 1430 985" type="button" value="?"/>
(3) $\text{ACH}^-$	<input type="radio"/> $\text{CH}_3\text{CO}$	(14) $\text{CNH}_2^-$		
<input checked="" type="radio"/> 6 $\text{ACH}$	<input type="radio"/> $\text{CH}_2\text{CO}$	<input type="radio"/> $\text{CH}_3\text{NH}_2$		
<input type="radio"/> $\text{AC}$		<input type="radio"/> $\text{CH}_2\text{NH}_2$		
		<input type="radio"/> $\text{CHNH}_2$		

Figure 19. UNIFAC module card used to select UNIFAC groups.

This card is connected to two other similar cards showing the rest of the available groups. To select a group, the user clicks on the group symbol and a dialog box will then appear to enable the user to input the number of this group that is in the molecule. The user then continues to select groups until all the groups that are in the molecule have been selected. When the user clicks the "Done" button the UNIFAC input string is built and returned to the UNIFAC card. In addition, for users familiar with the UNIFAC approach, the appropriate subgroups can also be entered directly by simply typing the number of a group followed by a space then the symbol of the group for each group in the chemical.

The final form of the input is "# group # group ....". The # represents the number of the functional group in the molecule, the group is the group symbol of the functional group. For example the UNIFAC input string for Toluene is "5 ACH 1 ACCH3", meaning five aromatic carbons with one hydrogen atom (5 ACH) and one aromatic carbon connected to a methyl group (1

ACCH3). The UNIFAC groups available to PEP are shown in Table 2. Remember UNIFAC subgroups may not be available for every compound. In these cases the activity coefficient can not be calculated using the UNIFAC.

Table 2. UNIFAC groups

Group Name	Group Symbol	Group Name	Group Symbol
CH2	CH3	CCN	CH3CN
	CH2		CH2CN
	CH	COOH	COOH
C=C	CH2=CH		HCOOH
	CH=CH	CCl	CH2Cl
	CH2=C		CHCl
	CH=C		CCl
ACH	ACH	CCl2	CH2Cl2
	AC		CHCl2
ACCH2	ACCH3	CCl3	CHCl3
	ACCH2		CCl3
	ACCH		CCl2
OH	OH	CCl4	CCl4
CH3OH	CH3OH	ACCl	ACCl
H2O	H2O	CNO2	CH3NO2
ACOH	ACOH		CH2NO2
CH2CO	CH3CO		CHNO2
	CH2CO	ACNO2	ACNO2
CHO	CHO	CS2	CS2
CCOO	CH3CO	CH3SH	CH3SH
	CH2COO		CH2SH
HCOO	HCOO	Furfural	Furfural
CH2O	CH3O	DOH	(CH2OH)2
	CH2O	I	I
	CH-O	Br	Br
	FCH2O	C=C	CH=C
CNH2	C(CH3)2NH2		C=C
	C(CH3)2NH2	DMSO	DMSO
	CHNH2	ACRY	ACRY
CNH	CH3NH	ClCC	Cl-(C=C)
	CH2NH	ACF	ACF
	CHNH	DMF	DMF-1
(C)3N	CH3N		DMF-2
	CH2N	CF2	CF3
ACHN2	ACHN2		CF2
Pyridine	C5H5N		CF
	C5H4N		
	C5H3N		

## **Calculate Activity Coefficients**

After the functional groups are chosen, the activity coefficients can be calculated using the procedure described by Grain (1990) by clicking on the "Calc. Activity Coefficients" button. Once the activity coefficients have been calculated they can be displayed, along with values from several intermediate steps, by clicking on the "Display Act. Coeff." button. The equations used to calculate the activity coefficients can also be displayed by clicking the "balloon" on the "UNIFAC Calculations" card.

## **Editing Parameters**

The UNIFAC group values can be edited by clicking on the "Edit Parameters" button. This will take you to an Index card containing a button for each group. To edit the value for that group, click on the corresponding group button. To edit the Q and R values, click on the left arrow at the bottom of the index card.

## **Estimating Properties**

To estimate S and/or  $K_{OW}$  click on the "Estimate Properties" button. Clicking the "eye" next to the property will display the equations used to estimate S or  $K_{OW}$ . The results will be reported on the results card.

## **Limitations of UNIFAC Approach to Estimating S and $K_{OW}$**

The use of UNIFAC is limited to chemicals that have subgroups contained in a UNIFAC data base. If a chemical has subgroups for which no UNIFAC values are available, the activity coefficients cannot be calculated and, therefore, the properties cannot be estimated. In addition, it has been observed that the errors associated with estimating S and  $K_{OW}$  via the UNIFAC approach tend to increase as the compound becomes less solubility (larger  $K_{OW}$ ). Correction factors have been presented in the literature to correct for this tendency.

## Property/Property Module

### Overview

Quantitative Property-Property Relationships (QPPRs), based on the relationship between two properties as determined by regression analysis, are used to predict the property of interest from another more easily obtained property without a specific concern for molecular structure. Frequently, the regression expressions are expressed in terms of the log of the two properties.

The operation of the Property-Property correlation module, illustrated in Figure 20, is considerably different than the other modules because the choice of the regression models used for the property estimation determines the required inputs.

**PEP Processor**

File Edit Go Print View

**Property/Property**

Chemical Name: 2,2',6,6'-tetrachlorobiphenyl

SMILES String:

**1. Choose Property**

- ☒ S
- ☐ Kow
- ☒ Koc
- ☒ BCF
- ☐ H

**2.**

☒ USER choose regression  
☐ PEP choose regression

**View Stats**

- Universal from Kow
- Universal from S & Kow
- Universal from Kow
- Universal from Kow
- PUS-RT

**3. Input Properties**

Look in Prop. DB

log S:

log Kow: 5.936

log Pv:

log H:

log Koc:

log EX:

M.P.:  °C

M.W.:

**4. Estimate Properties**

Note: All values are at 25°C

Figure 20. Example card from PEP property-property module.

## **Selecting the Properties to be Estimated**

Since the input for the Property/Property module depends on the properties to be estimated and the regression models used, the user must first select the properties to be estimated and the regression equations to be used before any input values are requested. The program keeps track of the properties required and provides the appropriate input fields. If available, the required properties can be imported directly from the associated chemical property database. Currently property-property relationships for the estimation of S, Kow, Koc, and BCF are available and will darken as the properties are selected.

## **Choosing the Property-Property Regression Model**

The most appropriate model to use in estimating the property of interest depends on the class of the chemical and what property-property relationships are available. The property-property relationships that are available in PEP are divided into two general categories; regressions developed using data from the PEP Chemical Physical Data Base and regressions obtained from the literature. The two categories are separated by a grey line in the popup menu. To choose the property-property model, use the popup menu directly under the current regression by holding the mouse button down while over the title of the current regression. The "eye" icon to the right of the regression title will display the regression equation and statistics associated with the current relationship. The "book" icon will show the reference for the current regression if the regression is from the literature.

## **Viewing the Calculated Values**

Once the properties and the relationships are chosen, the properties required for input need to be entered. The properties that are required are shown on the right side of the card with a line next to the symbol. The values can be retrieved from the PEP Chemical Property Database by making sure the property name is correct and clicking on the "Look in Prop. DB" button. The current units of the inputs will be shown if appropriate, and will be changed as needed by PEP for the different regression equations.

When the "Estimate Properties" button is clicked the properties that were input will be used along with the chosen regression equations to estimate the selected properties. If the values for the input properties have not been entered on the Property-Property card the user will be prompted to supply the required values. The property estimates along with the calculated 95% prediction interval, if the statistical information is available, will be placed on the Results card.



## **Limitations of the Property/Property Module**

One major limitation to the use of property/property relationships is the lack of suitable property values required for input into the regression models. For many new chemicals, no physical/chemical property information is available. In addition, the quality of the property-property regression model depends on the extent of the database used to develop the relationship. For example, some property-property models are broader than others in the type and number of compounds used in its development.

# PEP BATCH

## Overview

PEP Batch provides users with a method for the convenient, unattended calculation of MCIs, TSA and UNIFAC activity coefficients and subsequent estimation of physical properties for large numbers of compounds via the PEP processor described earlier. Like the PEP Processor, PEP Batch is divided into MCI, TSA and UNIFAC modules. Each module, as illustrated in Figure 21 for the MCI module, requires the user to select the appropriate input file (i.e. SMILES string, connection table or 3-D atomic coordinates), choose to information to be sent to the output file (i.e. chemical name, SMILES string, properties) and start the batch driver.

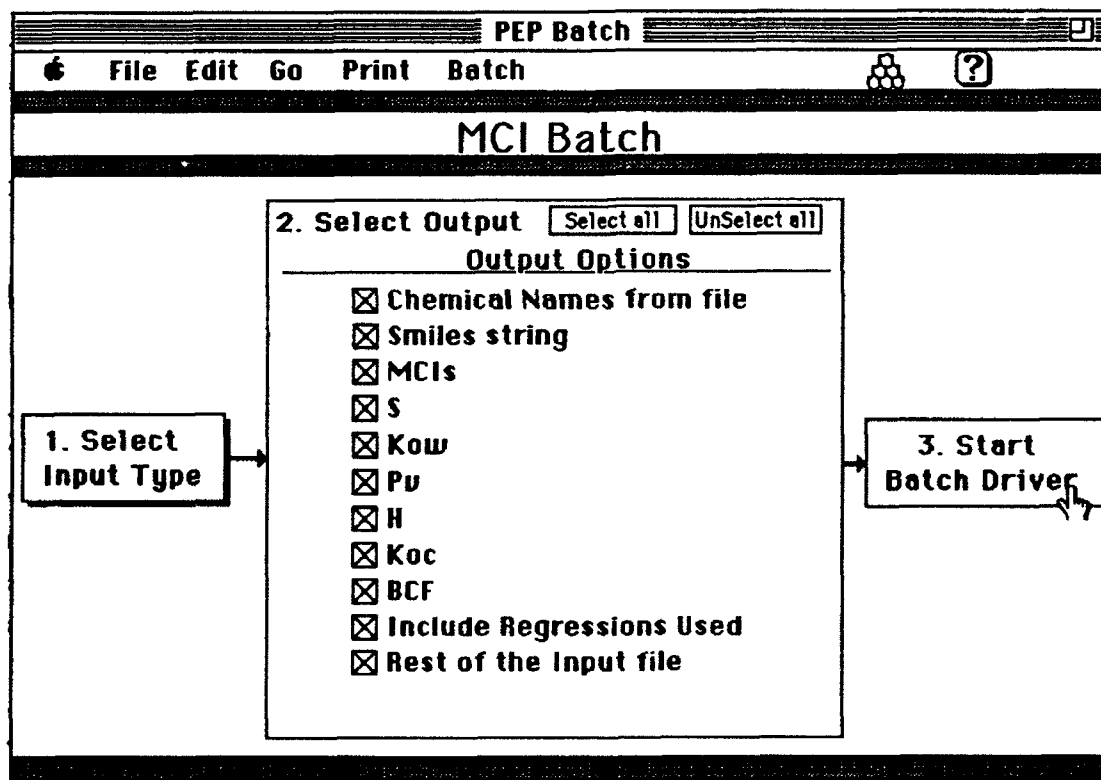


Figure 21. Example card for PEP Batch, MCI module.

## Input Structure

The MCI and UNIFAC modules require the two-dimensional molecular structure of the chemical to be entered using either SMILES strings or connection tables.

To enter SMILES strings into the MCI or UNIFAC batch modules, you must create a text file containing name of the chemical and its corresponding SMILES strings in columns separated by tabs and hard returns at the end of each row. The text file can contain additional tab-delimited information, but the SMILES strings and chemical names must be in the first or second column. When you select the SMILES input option, you must indicate the column order on the "Explain SMILES file card shown in Figure 22. The text file can be created with word processing or spreadsheet programs or you can also edit or create a file containing the SMILES strings and chemical names within PEP by selecting SMILES from the "input structure type" pulldown menu.

PEP Batch

File Edit Go Print Batch

Explain SMILES file

View File Contents

benzene c1ccccc1  
toluene c1cc(C)ccc1  
o-xylene c1c(C)c(C)ccc1  
ethylbenzene c1c(CC)ccc1

Column 1	Column 2
<input checked="" type="radio"/> Chemical Name	<input type="radio"/> Chemical Name
<input type="radio"/> SMILES	<input checked="" type="radio"/> SMILES

☐ Ignore First Line

Figure 22. PEP Batch, explain SMILES file card.

To input connection table files into the MCI or UNIFAC batch modules, you must first place them in a single folder. Select the "Connection tables" option from the "Select Input Type" popup button and choose the folder that contains the connection table files using the standard "open file" dialog box that appears. Highlighting any one of the files in the folder selects all of the files in that folder and allows you to view or delete specific files. The files can be any valid type for the MCI module as they will be converted if possible (see MCI module helps). After selecting the input files, click on the advance arrow located at the upper right of the card to advance to the next step.

The TSA batch module operates in the same manner as the MCI and UNIFAC modules except that the calculation of TSA requires the three-dimensional cartesian coordinates for each atom in the chemical of interest. The TSA batch module accepts Cartesian Coordinates or Alchemy files. Alchemy files contain both the two-dimensional chemical structure and the coordinates. This allows PEP to calculate the chemical's TSA and determine the most appropriate TSA-property relationship based on chemical class. From within the standard dialog box, you can click on any file in a folder to select all of the files in that folder. The files will then be displayed. You can also view or delete files.

### **Output Options**

Once the input files have been selected, the "output option" step becomes active. This allows the user to select the properties to be calculated and any additional information that is available (i.e. chemical name, SMILES string, MCIs, TSA, or UNIFAC activity coefficients) to be exported by to a tab-delimited text file.

### **Start Batch Driver**

After the output information is selected, the "Start Batch Driver" button becomes active. Clicking this button brings up the standard Macintosh "save file" dialog box that allows the user to specify the name of the data file to be exported by PEP batch and the location that the file will be saved.

---

## ***CHEMICAL PROPERTY DATABASE***

### **Overview**

Experimentally determined physical property data for about 800 compounds, having at least one value of aqueous solubility (S), octanol/water partition coefficient (Kow), vapor pressure (Pv), organic carbon normalized soil sorption coefficient (Koc), bioconcentration factor (BCF), or Henry's law constant (H), was compiled from a variety of literature sources and computerized databases. Using this information, a chemical property database was constructed using HyperCard™ and subsequently used for developing MCI-property, TSA-property and property-property relationships. In addition to the properties listed above, the database includes the following information: compound name and synonyms, a diagram of the 2-D chemical structure, SMILES notation, uses, CAS number, chemical formula, molecular weight (MW), boiling point (BP), melting point (MP), and appropriate references for each value. A built-in unit conversion utility enables users to quickly view property values in a variety of commonly used units. The database is directly connected to the PEP Processor stack.

The Chemical Property Database also provides the means for the user to search for chemical compounds by full or partial name or synonym, to sort the compounds by name, boiling point, melting point, or molecular weight, and the ability to transfer to any of the property estimation modules. In addition, the user can easily edit existing values, add new values or export information to a text file or another database.

In addition to the physical properties, information describing the environmental persistence and toxicity of specific chemicals can also be entered into the database. Placeholders for biodegradation rates, hydrolysis rates, photolysis rates and LC50s, along with the appropriate references and comments have been incorporated into the database. This feature was added to the database after requests from test users, however currently no degradation or toxicity data has been entered into the database. The chemical property database is illustrated in Figures 23 and 24.

**Chemical Property Data Base**

File Edit Go Find Sort Values Print Export ? 179

**2,2',6,6'-tetrachlorobiphenyl**

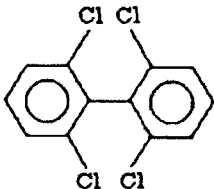
CAS #: 15968-05-5

Formula: MW: 292

MP: 200 °C BP: °C

Synonyms:

View: **Structure**



View: **Chemical\Physical Properties**

log/antilog	Physical/Chemical Properties	Values	Temp °C	Units	Ref
log	Aqueous Solubility (S)				
log	Octanol/Water Partition Coefficients (Kow)	5.936	25		
log	Vapor Pressure (Pv)				
log	Henry's Law Constant (H)	-1.648	25	dimensionless	
log	Soil/Water Sorption Coefficient (oc-based) (Koc)				
log	Bioconcentration Factor (BCF)				
log	Acid Dissociation (Ka)				

Figure 23 Example card from the Chemical Property Data Base

**chemical property data base**

File Edit Go Find Sort Values Print Export ? 179

**2,2',6,6'-tetrachlorobiphenyl**

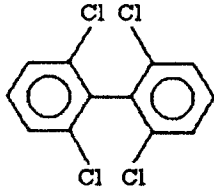
CAS #: 15968-05-5

Formula: MW: 292

MP: 200 °C BP: °C

Synonyms:

View: **Structure**



View: **Chemical\Physical Properties**

**Degradation Properties**

Degradation	t 1/2	Matrix	Temp (°C)	Comments
Hydrolysis				
Photolysis				
Biodegradation				

Figure 24. PEP database degradation properties.

## Searching for Chemical Compounds in PEP's Database

Chemicals in the database can be found by full chemical name, partial chemical name, synonyms, or CAS number searches, or by selecting the chemical from a current list of compounds. The user selects the desired option from the "Find" menu.

Searching or Finding by either chemical name, partial chemical name, synonyms or CAS number is done using the finding tools built into HyperCard. When the finding option is chosen a standard dialog box is used to input the string that is to be found. Finding by full chemical name or CAS number requires that the entire contents of the appropriate field match the string of characters that was entered. Finding by partial chemical name or synonym requires that all of the string that was entered be present in the appropriate field.

The select from list option allows the user to select a compound from the current list of chemical names. The chemical names appear alphabetically sorted in a dialog box. Typing the first letter or number of the chemical will scroll the list to the first chemical starting with that character. The chemical card is located and displayed by clicking twice on the same chemical name or selecting the chemical name and then clicking the "Go There" button.

## Sorting the Database

The "Sort" menu at the top of the card allows you to sort the database by name, melting point (MP), boiling point (BP), or molecular weight (MW). NOTE: the alphabetic sort function built into HyperCard places all chemicals with numbers preceding the letters in the chemical name before those chemicals having only letters in the chemical name. For example, after sorting by name, 1-octanol would come before acetone and 2-octanol, but after 1-decanol.

## Adding or Deleting Data

New chemicals may be added to the data base by choosing "Add new card" option from the Values menu. The user will then be prompted for the chemical name, CAS number, molecular formula, melting point, boiling point, molecular weight, and the SMILES string.

Additional values for melting point, boiling point, S,  $K_{ow}$ ,  $P_v$ , H,  $K_{oc}$ , BCF, and  $K_a$  can be added to a chemical already in the database using the "Add value" option accessed from the Value menu. As each new value is entered the user is prompted for the reference. The reference can be an existing reference or the user can add a new reference. If a new reference is to be added, the user will be prompted for the author, year, title, publication, volume, issue, and the page numbers. The user will also be given a chance to edit the new reference.

An entire card or specific values contained on a card can be deleted by choosing the appropriate item from the Value menu. A reference can also be deleted if it was entered incorrectly. The data will not be deleted until the user verifies the request through a standard dialog box.

### **Printing Information From the Database**

The user can print the current card, or a series of cards using the print features built into HyperCard. To print a series of cards, first the user must select that option from the "Print" menu. Then while each card that is to be printed is the currently displayed card, the select "Print this card" option from the print menu must be chosen. When the desired cards have been printed, the user then selects "Stop printing" from the "Print" menu. The user can also print a list of all the references contained in the PEP data base.

### **Exporting Information From the Database**

Information in the database can either be exported as text in a line or as a spreadsheet format using the "Export" menu. If the line format option is used, the data will appear with each data field on a line and each chemical separated by a solid line. This option is most appropriate if a word processor will be used to print or manipulate the exported data. If the spreadsheet format is used the data will be put in columns delimited by a tab.

### **Moving to Other PEP Modules**

The "Go" menu at the top of the card allows you to move to any PEP module. To use the Go option, move the pointer to Go menu at the top of the card. Hold the pointer down on the button until the selections appear. Then move the pointer until the selection of choice is highlighted and release the mouse button.

### **Changing the Units of Measurement**

The popup menu under each property's units allows you to change the units for solubility, Henry's law, and vapor pressure. All values of the property will be changed to reflect the new units.



## PEP Models

### Overview

To illustrate the practical application of PEP, an additional stack called PEP Models was developed. This stack, which contains the algorithms for the Level 1 and 2 Fugacity Models (Mackay, 1980), is linked directly to the the PEP Processor, but can also be used independently. The Level 1 Fugacity Model considers a unit world consisting of six compartments: air, water, soil, suspended solids, sediment, and biota (as shown in Figure 25).

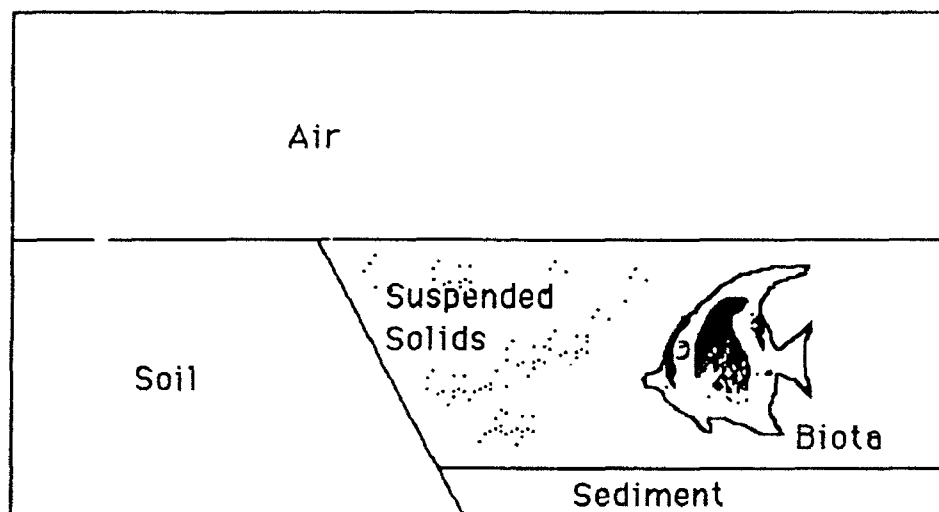


Figure 25. Representation of Fugacity Level 1 compartments.

The model predicts the equilibrium concentrations of the chemical of interest in each compartment using the fugacity approach described by Mackay (1980). The model requires the input of  $K_{OC}$ ,  $H$  and  $BCF$  which can be read directly from the PEP processor or the PEP chemical property database, if available. In addition to the chemical specific properties, the density and volume of each compartment must be specified along with the organic carbon content of the soil, sediment and suspended sediment. An editable set of default values for compartment density, volume and organic carbon content, as shown in Figure 26.

**PEP Models**

**File Edit Go Print Compartments**

## Fugacity Model

**Chemical Name: 2,2',6,6'-tetrachlorobiphenyl**

**1. Input Property Values**

Look for Values in Prop.DB

log Koc ☐ 3.92

log H ☒ (dimensionless) -1.648

log BCF ☐ 4.69

☒ Indicates value from DB

**2. Input Environmental Compartment Values**

Compartment	Density kg/m <sup>3</sup>	Volume m <sup>3</sup>	% Organic Carbon
<input checked="" type="checkbox"/> Air	1.19	1e10	
<input checked="" type="checkbox"/> Water	1000	7e6	
<input checked="" type="checkbox"/> Soil	1500	9e3	2
<input checked="" type="checkbox"/> Susp. Solids	1500	35	4
<input checked="" type="checkbox"/> Sediment	1500	2.1e4	4
<input checked="" type="checkbox"/> Biota	1000	3.5	

**3. Calculate Distribution**

OR

**Input Fugacity Level 2 Data**

Figure 26. PEP Models card, input for Fugacity Level 1

The Level 2 Fugacity Model allows for the chemical of interest to degrade in each compartment, move by advection through the water and air phases, and be emitted into the unit world. The rate values for each of these processes must be entered by the user. The degradation rates for each compartment can be entered either by  $t_{1/2}$  values in hours or by first order reaction rate constants in 1/hours, as shown in Figure 27. The advection rate data can be entered either by residence time or flow rate and the concentration or by directly entering the mass flow rate in moles per hour. The emission rate is entered in the units of moles per hour.

**PEP Models**

**File Edit Go Print Rates**

**Fugacity Level 2**

**Chemical Name: 2,2',6,6'-tetrachlorobiphenyl**

**3. Input Reaction Rate Data**  
(input reaction 1/2 life or reaction rate constant)

Reaction 1/2 life hours	OR	Reaction rate const. hours <sup>-1</sup>
_____		_____
_____		_____
_____		_____
_____		_____
_____		_____
_____		_____

**4. Input Advective Data**  
(input mass flow rate or conc. and residence time or flow rate)

Residence time hours	OR	Flow Rate m <sup>3</sup> /hr	Conc. mol/m <sup>3</sup>	Mass flow rate mol/hr
_____		_____	_____	0.0E+00
_____		_____	_____	0.0E+00

☒ Air

☒ Water

☒ Soil

☒ Susp. Solids

☒ Sediment

☒ Biota

**5. Input Emission Rate**  
Emission rate mol/hr \_\_\_\_\_

↓

Input Level 3 Data

OR

6. Calculate Distribution

Figure 27. PEP Models card, input for Fugacity Level 2

### Input Property Values

To calculate the environmental distribution of a chemical the user must enter the several physical-chemical properties along with the size and densities of each compartment. The physical-chemical properties that are need to be input for the six compartment unit world are Koc, H, and BCF. However, in step 2 the user may choose to eliminate one or more compartments to simplify the model or to more closely examine the distribution of a chemical in a sub-environment such as the hypolimnion of a lake (i.e., water, air, biota). This would eliminate the need for a value for Koc. PEP determines which physical-chemical property values are needed depending on the unit world which the user has defined. The chemical-physical properties that are not needed will be dimmed.

When available, experimental values are preferred over estimated values in environmental fate modeling. Experimental values for the physical-chemical properties can be searched for in the Chemical Property Data Base by using the "Look for Values in Prop.DB" option. When clicked, a

full chemical name search is initiated, and if the chemical is found, the values of the three properties required by the Fugacity level 1 model will be displayed in a dialog box. The values can be copied to the Fugacity card using the buttons on the dialog box as illustrated in Figure 28. The check boxes next to the property names of the Fugacity card indicate if the property came from the data base.

**Values from the PEP Property Database**  
**Select the properties to copy to the Model**

Log Koc : 1.780
Log H : -0.648
Log BCF : Not Found

Buttons: Cancel, Copy All, Copy Selected

Figure 28. Example dialog box resulting from the "Look for Values in Prop.DB" option

### Input Environmental Values

The values used for the unit world can be changed in step 2. The available compartments are air, water, soil, suspended solids, sediment, and biota. Water must always be part of the unit world because of the partition coefficients are defined as the ratio of the chemical in the non-aqueous phase to the aqueous phase. The rest of the compartments can be eliminated by clicking the check box next the compartment name. The densities, volumes, and the percent organic carbon (if appropriate) of each compartment can be changed.

Different user defined environments or unit worlds can also be saved using the "Save Current Values" button and later recalled using the "Values:" popup menu. The "Delete Saved Values" button can be used to remove a saved set of values from the list that will popup when the "Values" button is used. The "Set Default Values" button can be used to set the unit world that is used when the PEP Models stack is opened.

### Calculate Distribution

After the user inputs all the required information and hits the "calculate distribution" button, the model calculations are performed in HyperTalk and the results are presented in both in tabular and

graphical form as illustrated in Figure 29. The graphical display can be changed from bar (concentration of the chemical in each phase) to pie (percent of the chemical in each phase) chart forms using the "Graph:" popup menu. The values of the distribution coefficients that were used in the calculations are also shown on the results card.

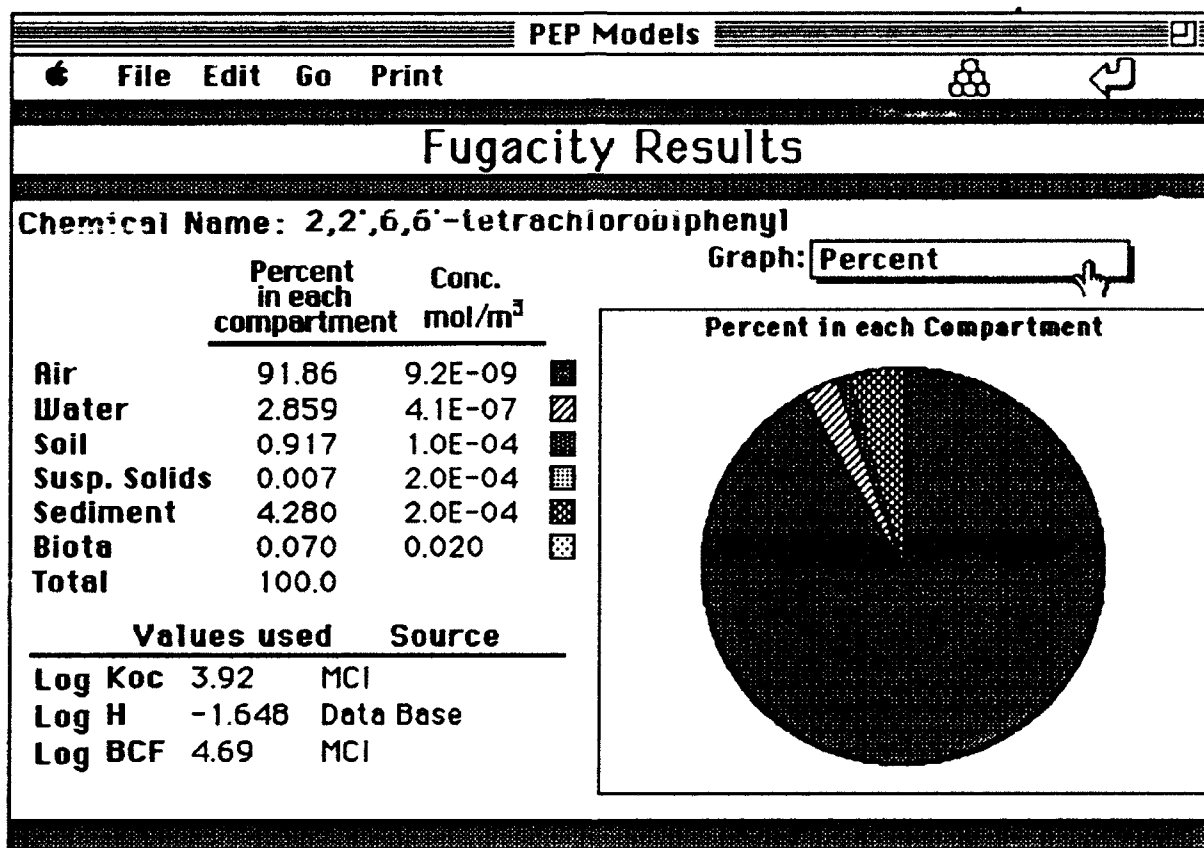


Figure 29. PEP Models results card

## PEP Help

Information regarding the operation of the chemical property database and the property estimation, models and batch modules is available in the PEP Help stack. This stack easily accessed at any time within the PEP system. The organization and layout of each help card is similar to that illustrated in Figure 30 for the MCI module. The user can select the topic of interest by clicking on the appropriate radio button and the information on that subject will be displayed in the scrolling field.

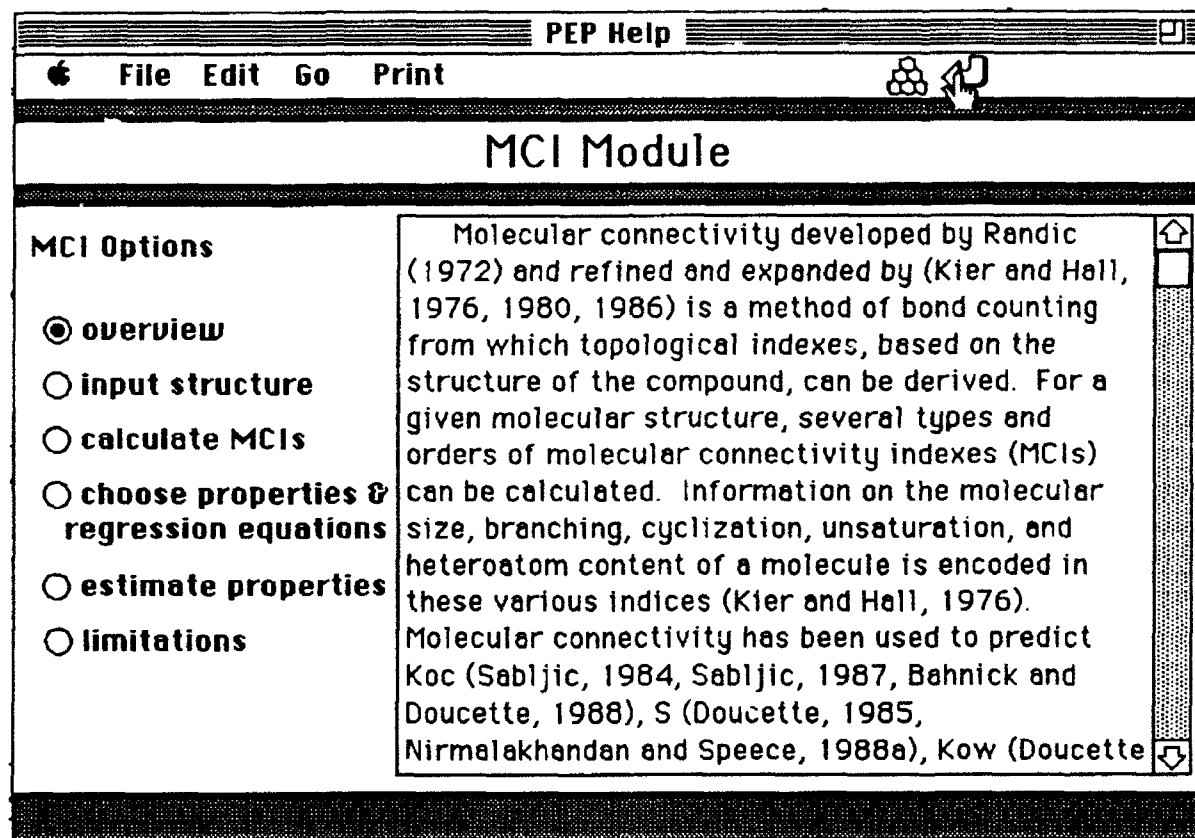


Figure 30. Example card from PEP Help stack for MCI module.

---

## REFERENCES

- Anderson, E., G.D. Veith, and D. Weininger. 1987. SMILES: A Line Notation and Computerized Interpreter for Chemical Structures. EPA/600/M-87/021.
- Arbuckle, W.B. 1986. Using UNIFAC to Calculate Aqueous Solubilities. *Environ. Sci. Technol.*, 20: 1060-1064.
- Bahnick, D.A. and W.J. Doucette. 1988. Use of Molecular Connectivity Indices to Estimate Soil Sorption Coefficients for Organic Chemicals. *Chemosphere*. 17(9): 1703-1715
- Banerjee, S. 1985. Calculation of Water Solubility of Organic Compounds with UNIFAC-Derived Parameters. *Environ. Sci. Technol.* 19(4): 369-370.
- Banerjee, S. and P.H. Howard. 1988. Improved Estimation of solubility and Partitioning Through Correction of UNIFAC-Derived Activity Coefficients. *Environ. Sci. Technol.* 22(7): 839-841.
- Cambridge Scientific. 1989. ChemDraw and Chem3D+. Cambridge, Massachusetts [Computer Programs]
- Gmehling, J., P. Rasmussen, and A. Fredenslund. 1982. Vapor-liquid equilibria by group contribution. Revision and Extension. 2. *Ind. Eng. Chem. Process Des. Dev.* . 21: 118-127.
- Grain, C. F. 1990. Activity Coefficient, Ch. 11. In Lyman, W.J. W.F. Reehl and D.H. Rosenblatt (Eds.). *Handbook of chemical and property estimation methods*. American Chemical Society. Washington, D. C .
- Kier, L.B., Hall, and L. H., *Molecular Connectivity in Structure-Activity Analysis*. 1986, Letchworth, Hertfordshire, England.: Research Studies Press, LTD., England, and John Wiley & Sons, Inc., New York. 262.
- Mackay, D. 1979. Finding fugacity feasible. *Environ. Sci. Technol.* 13(12): 1218-1223.
- Mackay, D. and S. Paterson. 1981. Calculating Fugacity. *Environ. Sci. Technol.* 15(9): 1006-1014.

- Pauling, L. 1960. The nature of the chemical bond. Cornell University Press, Ithaca, New York 262 p.
- Pearlman, R.S. 1980. Molecular surface areas and volumes and their use in structure/activity relationships, p. 331-346. In S.H. Yalkowsky, A.A. Sikula and S.C. Valvani (Eds.). Physical chemical properties of drugs. Marcel Dekker, Inc., New York.
- Tripos Associates Inc. 1989. Alchemy and CONCORD. St. Louis, Missouri [Computer Programs]
- Weininger, David. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28:31-36.



## **APPENDIX B**

### **STATISTICAL SUMMARY FOR MCI-PROPERTY, TSA-PROPERTY AND PROPERTY-PROPERTY RELATIONSHIPS INCORPORATED INTO PEP**

## STATISTICS Class: s Universal

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.3917	0.1378	2.85
vp1	-0.9257	0.0318	-29.3
avp1	1.8251	0.1047	17.4
MP-25	-0.01		

## Analysis of Variance Table

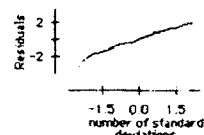
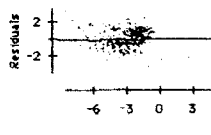
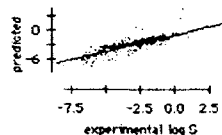
Source	RSS	df	MSS	F
Regression	889.176	2	444.588	44.6
Residual	360.920	362	0.997	
Total	1250.096	364	3.4343	

$r^2 = 71.1\%$   $\text{Obs} = 365$   $S = 0.9985$

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: s Ureas

## Regression Results

Variable	Coef.	Std. Error	t
Constant	3.758	0.4631	8.12
bp1	-1.0747	0.0808	-13.3
MP-25	-0.01		

## Analysis of Variance Table

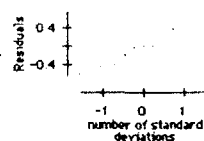
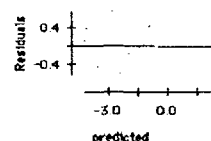
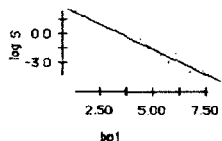
Source	RSS	df	MSS	F
Regression	32.5444	1	32.54	177
Residual	1.05604	9	0.1184	
Total	34.20044	10	3.42	

$r^2 = 95.2\%$   $\text{Obs} = 11$   $S = 0.4290$

Log S vs. MCI

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: s Alcohols

## Regression Results

Variable	Coef.	Std. Error	t
Constant	2.8524	0.0839	34
bp1	-1.1446	0.0201	-57
MP-25	-0.01		

## Analysis of Variance Table

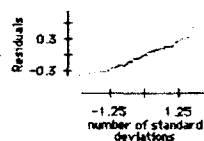
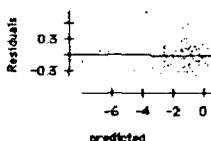
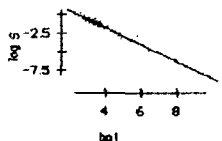
Source	RSS	df	MSS	F
Regression	167.948	1	167.948	3248
Residual	3.41245	58	0.0588	
Total	171.3604	67	2.5576	

$r^2 = 98.0\%$   $\text{Obs} = 68$   $S = 0.2274$

Log S vs. MCI

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Pv Universal

## Regression Results

Variable	Coef.	Std. Error	t
Constant	6.2811	0.1801	29.2
np3	-1.2748	0.0491	-25.9

## Analysis of Variance Table

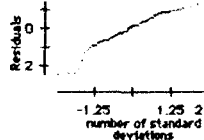
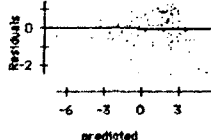
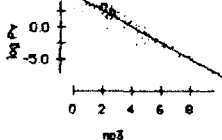
Source	RSS	df	MSS	F
Regression	453.998	1	453.998	672
Residual	62.7848	93	0.675	
Total	516.7828	94	5.4977	

$r^2 = 87.9\%$   $\text{Obs} = 95$   $S = 0.8218$

Log Pv vs. MCI

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: s PCBs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.8528	0.2204	-8.41
bp6	6.2017	0.3654	17
MP-25	-0.01		

## Analysis of Variance Table

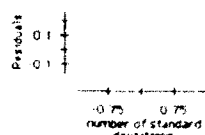
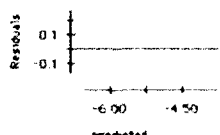
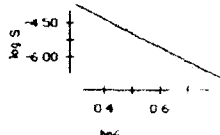
Source	RSS	df	MSS	F
Regression	6.50320	1	6.5032	288
Residual	0.158026	7	0.0226	
Total	6.661226	8	0.8327	

$r^2 = 97.6\%$   $\text{Obs} = 9$   $S = 0.1503$

Log S vs. MCI

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: s Anilines

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.9892	0.401	2.47
bp1	-0.7157	0.0734	-9.75
MP-25	-0.01		

## Analysis of Variance Table

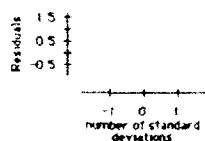
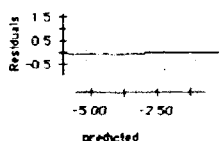
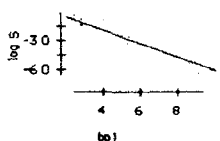
Source	RSS	df	MSS	F
Regression	45.5658	1	45.57	95.0
Residual	3.10898	19	0.1626	
Total	54.67478	20	2.7337	

$r^2 = 83.3\%$   $\text{Obs} = 21$   $S = 0.6924$

Log S vs. MCI

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: s PAHs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.1353	0.2894	-3.92
np1	-0.5011	0.0373	-13.4
MP-25	-0.01		

## Analysis of Variance Table

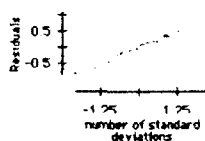
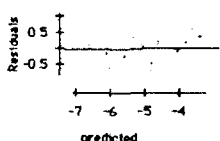
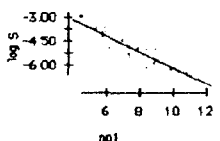
Source	RSS	df	MSS	F
Regression	30.5352	1	30.54	181
Residual	5.74826	34	0.1691	
Total	36.28346	35	1.0367	

$r^2 = 84.2\%$   $\text{Obs} = 36$   $S = 0.4112$

Log S vs. MCI

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Pv PCBs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	9.4791	1.815	5.22
np3	-2.4284	0.4481	-5.42
nc5	6.8143	2.15	2.7

## Analysis of Variance Table

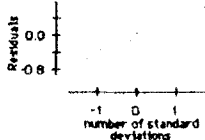
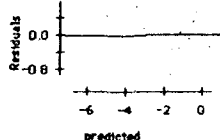
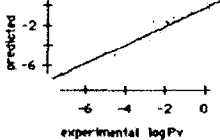
Source	RSS	df	MSS	F
Regression	49.8285	2	24.91	115
Residual	2.60424	12	0.217	
Total	52.43274	14	3.7452	

$r^2 = 95.0\%$   $\text{Obs} = 15$   $S = 0.4659$

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Koc Universal

## Regression Results

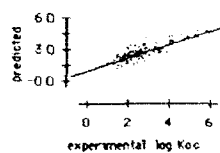
Variable	Coef.	Std. Error	t
Constant	0.6111	0.1767	3.46
np1	0.4647	0.0289	16.1
avp0	-1.2243	0.1176	10.4

## Analysis of Variance Table

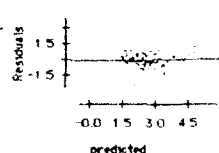
Source	RSS	df	MSS	F
Regression	159.141	2	79.57	138
Residual	98.8154	19	5.201	
Total	257.9564	21		

 $r^2 = 62.2\%$   $n_{obs} = 21$   $S = 0.7583$ 

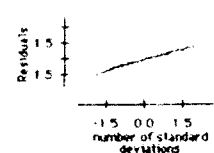
### Predicted vs. Exp.



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc Triazines

## Regression Results

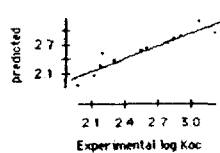
Variable	Coef.	Std. Error	t
Constant	-5.843	0.7221	-7.81
bp1	0.6296	0.0919	6.85
avp0	-0.9382	0.3763	-2.49

## Analysis of Variance Table

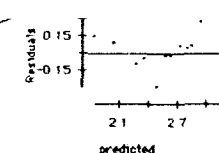
Source	RSS	df	MSS	F
Regression	1.60125	2	0.8006	32.8
Residual	0.221318	9	0.0246	
Total	1.822568	11		

 $r^2 = 87.9\%$   $n_{obs} = 12$   $S = 0.1568$ 

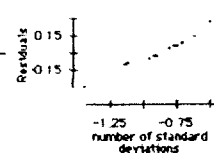
### Predicted vs. Exp.



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc PAHs

## Regression Results

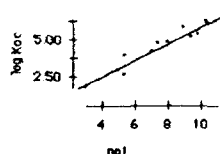
Variable	Coef.	Std. Error	t
Constant	0.2187	0.322	0.679
np1	0.5872	0.0405	14

## Analysis of Variance Table

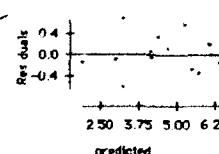
Source	RSS	df	MSS	F
Regression	24.1483	1	24.15	198
Residual	1.60011	13	0.1231	
Total	25.74841	14		

 $r^2 = 93.8\%$   $n_{obs} = 15$   $S = 0.3508$ 

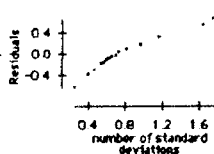
### Log Koc vs. MCI



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc Carbamates

## Regression Results

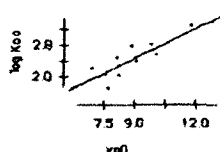
Variable	Coef.	Std. Error	t
Constant	0.1814	0.3295	0.551
vp0	0.2523	0.0363	6.96

## Analysis of Variance Table

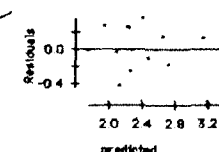
Source	RSS	df	MSS	F
Regression	2.83502	1	2.835	48.4
Residual	0.598675	11	0.0544	
Total	3.433695	12		

 $r^2 = 81.5\%$   $n_{obs} = 13$   $S = 0.2333$ 

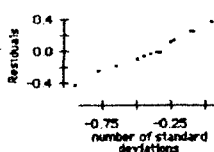
### Log Koc vs. MCI



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc Ureas

## Regression Results

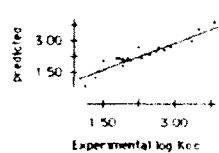
Variable	Coef.	Std. Error	t
Constant	0.8342	0.1578	5.29
np6	1.8588	0.148	11.2
nc6	-0.2556	1.511	-0.13

## Analysis of Variance Table

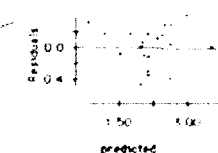
Source	RSS	df	MSS	F
Regression	9.29529	2	4.6476	68.8
Residual	1.28144	19	0.0675	
Total	10.57673	21		

 $r^2 = 67.9\%$   $n_{obs} = 22$   $S = 0.2599$ 

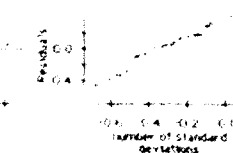
### Predicted vs. Exp.



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc PCBs

## Regression Results

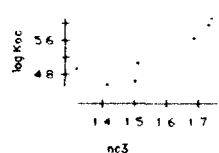
Variable	Coef.	Std. Error	t
Constant	0.0018	1.252	0.001
nc3	3.3877	0.799	4.21

## Analysis of Variance Table

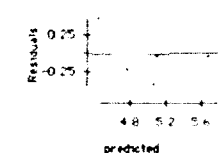
Source	RSS	df	MSS	F
Regression	1.84147	1	1.8415	17.8
Residual	0.518298	5	0.1037	
Total	2.359768	6		

 $r^2 = 78.0\%$   $n_{obs} = 7$   $S = 0.3220$ 

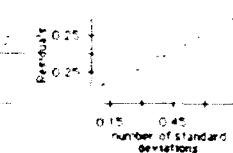
### Log Koc vs. MCI



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc Halogenated Aromatics

## Regression Results

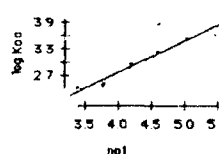
Variable	Coef.	Std. Error	t
Constant	-2.045	0.5708	-3.58
np1	0.7411	0.1318	5.62

## Analysis of Variance Table

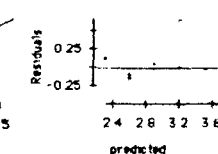
Source	RSS	df	MSS	F
Regression	2.03980	1	2.0398	31.6
Residual	0.515841	8	0.0645	
Total	2.555641	9		

 $r^2 = 79.8\%$   $n_{obs} = 10$   $S = 0.2539$ 

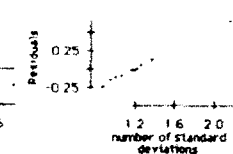
### Log Koc vs. MCI



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Koc Anilines

## Regression Results

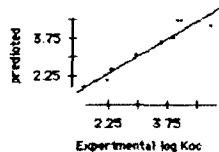
Variable	Coef.	Std. Error	t
Constant	0.1424	0.3348	0.425
np1	0.5734	0.0778	7.39
avp0	-1.2386	0.3295	-3.76

## Analysis of Variance Table

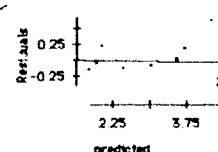
Source	RSS	df	MSS	F
Regression	9.94209	2	4.971	45.6
Residual	0.763154	7	0.109	
Total	10.70524	9		

 $r^2 = 92.9\%$   $n_{obs} = 10$   $S = 0.3302$ 

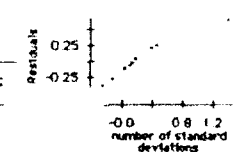
### Predicted vs. Exp.



### Residual vs. Predicted



### Residual vs. Prob.



# STATISTICS Class: Pv PAHs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	9.7188	0.4856	19.7
np1	1.4892	0.0702	21.2

## Analysis of Variance Table

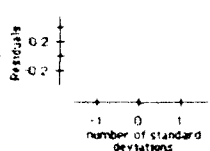
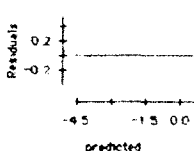
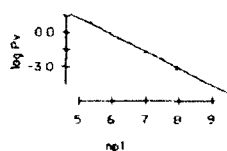
Source	RSS	df	MSS	F
Regression	35.5950	1	35.6	450
Residual	0.712338	9	0.0791	
Total	36.30733	10	3.6307	

 $r^2 = 98.0\%$   $n_{obs} = 11$   $S = 0.2813$ 

Log Pv vs MCI

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Universal

## Regression Results

Variable	Coef.	Std. Error	t
Constant	1.1527	0.1411	8.17
vp0	0.3978	0.0175	22.7
Avp0	-1.8843	0.1234	-16.1

## Analysis of Variance Table

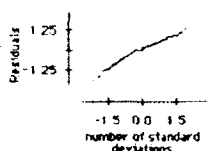
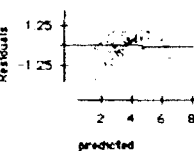
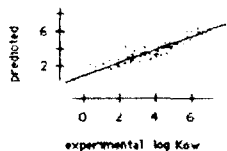
Source	RSS	df	MSS	F
Regression	335.264	2	168	315
Residual	102.809	193	0.5327	
Total	438.073	195	2.2485	

 $r^2 = 76.5\%$   $n_{obs} = 196$   $S = 0.7299$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow PCBs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	3.554	0.1888	18.8
bp4	0.9026	0.2587	3.52
bp6	5.8313	0.6189	9.42

## Analysis of Variance Table

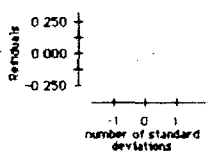
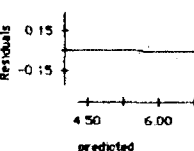
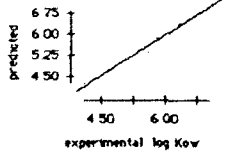
Source	RSS	df	MSS	F
Regression	9.80827	2	4.9031	16.7
Residual	0.382424	13	0.0294	
Total	10.18069	15	0.6782	

 $r^2 = 96.2\%$   $n_{obs} = 16$   $S = 0.1715$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Anilines

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.3184	0.2381	-5.54
bp1	1.0828	0.0878	12.3
Avp5	-3.8803	0.6307	-6.31

## Analysis of Variance Table

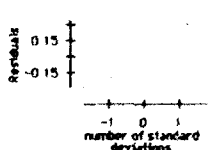
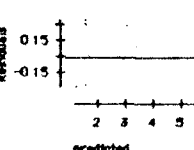
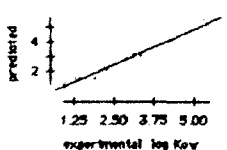
Source	RSS	df	MSS	F
Regression	18.5293	2	9.265	17.2
Residual	0.848843	12	0.0539	
Total	19.37814	14	1.3867	

 $r^2 = 96.8\%$   $n_{obs} = 15$   $S = 0.2322$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Pv Halogenated Aromatics

## Regression Results

Variable	Coef.	Std. Error	t
Constant	6.822	0.349	19
bp1	-1.5592	0.0701	-22.2

## Analysis of Variance Table

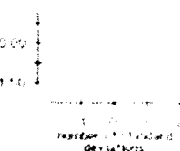
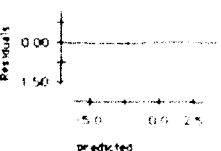
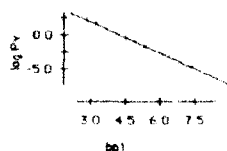
Source	RSS	df	MSS	F
Regression	147.169	1	147	494
Residual	7.14789	24	0.2978	
Total	154.3168	25	6.1727	

 $r^2 = 95.4\%$   $n_{obs} = 26$   $S = 0.5457$ 

Log Pv vs MCI

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Ureas

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-3.9333	0.2288	-17.2
vp1	1.2594	0.0581	22.5

## Analysis of Variance Table

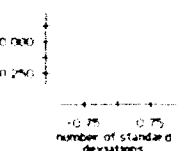
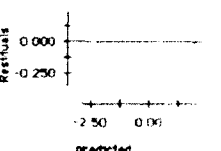
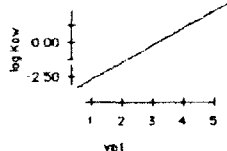
Source	RSS	df	MSS	F
Regression	16.8299	1	16.83	504
Residual	0.186679	5	0.0373	
Total	17.01657	6	2.8364	

 $r^2 = 99.0\%$   $n_{obs} = 7$   $S = 0.1932$ 

Log Kow vs MCI

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow PAHs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.7812	0.218	3.58
np0	0.4103	0.0208	19.9

## Analysis of Variance Table

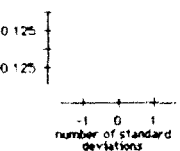
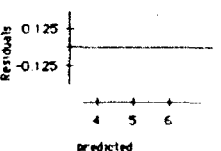
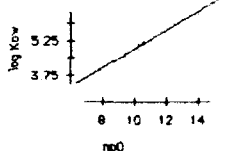
Source	RSS	df	MSS	F
Regression	10.8031	1	10.80	395
Residual	0.303592	11	0.0276	
Total	11.10669	12	0.9239	

 $r^2 = 97.3\%$   $n_{obs} = 13$   $S = 0.1651$ 

Log Kow vs MCI

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Alcohols

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.9813	0.1281	-15.3
bp1	1.1885	0.0364	32.9

## Analysis of Variance Table

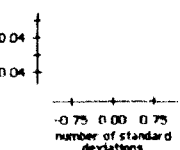
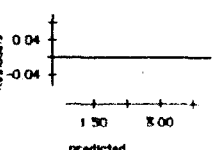
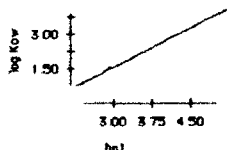
Source	RSS	df	MSS	F
Regression	5.03458	1	5.0348	1085
Residual	0.013924	3	0.0046	
Total	5.048484	4	1.2621	

 $r^2 = 99.7\%$   $n_{obs} = 5$   $S = 0.0681$ 

Log Kow vs MCI

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: H Universal

## Regression Results

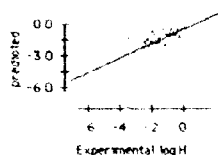
Variable	Coef	Std. Error	t
Constant	0.0003	0.2086	0.001
np1	-2.198	0.0422	-5.91
avp1	-1.667	0.2082	-8.01

## Analysis of Variance Table

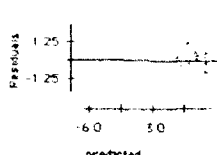
Source	RSS	df	MSS	F
Regression	93.9802	2	46.99	74.7
Residual	45.9296	73	0.6292	
Total	139.9098	75	1.8655	

 $r^2 = 67.2\%$   $N_{obs} = 76$   $S = 0.7932$ 

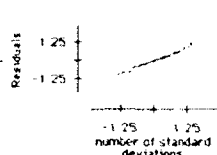
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: H PAHs

## Regression Results

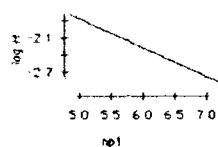
Variable	Coef	Std. Error	t
Constant	0.9541	0.2958	2.89
np1	5.197	0.0485	10.5

## Analysis of Variance Table

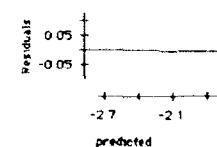
Source	RSS	df	MSS	F
Regression	0.687162	1	0.6872	110
Residual	0.024924	4	0.0062	
Total	0.712086	5	0.1424	

 $r^2 = 96.5\%$   $N_{obs} = 6$   $S = 0.0789$ 

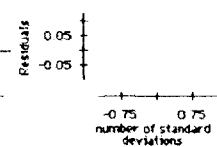
## Log H vs. MCI



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: BCF Universal

## Regression Results

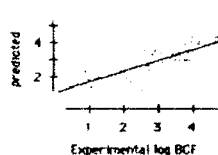
Variable	Coef	Std. Error	t
Constant	0.9816	0.2329	4.21
np1	0.4347	0.0415	10.5
avp0	-1.9999	0.2886	-6.93

## Analysis of Variance Table

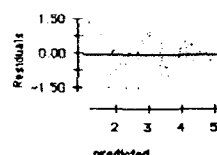
Source	RSS	df	MSS	F
Regression	60.1143	2	30.06	61.6
Residual	32.7178	67	0.4883	
Total	92.8321	69	1.3454	

 $r^2 = 64.6\%$   $N_{obs} = 70$   $S = 0.6988$ 

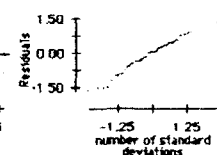
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: BCF Halogenated Aliphatics

## Regression Results

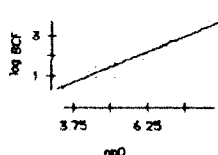
Variable	Coef	Std. Error	t
Constant	-1.7216	0.3818	-4.51
np0	0.632	0.0707	8.93

## Analysis of Variance Table

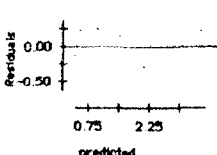
Source	RSS	df	MSS	F
Regression	8.51622	1	8.5162	79.8
Residual	0.853741	8	0.1067	
Total	9.369961	9	1.0411	

 $r^2 = 90.9\%$   $N_{obs} = 10$   $S = 0.3287$ 

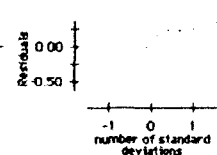
## Log BCF vs. MCI



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: Koc Sabljic 1988

# STATISTICS Class: H PCBs

## Regression Results

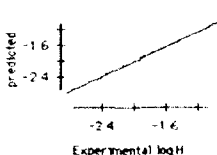
Variable	Coef	Std. Error	t
Constant	-2.477	0.099	-25
np6	0.3676	0.0247	14.9
vp4	1.1555	0.0895	12.9

## Analysis of Variance Table

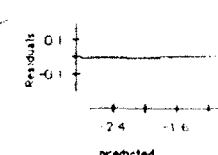
Source	RSS	df	MSS	F
Regression	3.26115	2	1.6306	111
Residual	0.175504	12	0.0146	
Total	3.436654	14	0.2455	

 $r^2 = 94.9\%$   $N_{obs} = 15$   $S = 0.1209$ 

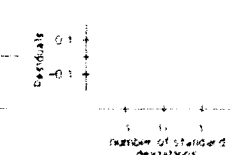
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: H Halogenated Aliphatics

## Regression Results

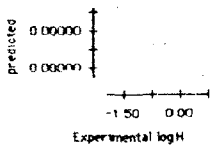
Variable	Coef	Std. Error	t
Constant	0.7325	0.2612	2.8
vp1	-1.2237	0.1559	-7.85
np2	0.6848	0.0961	7.13

## Analysis of Variance Table

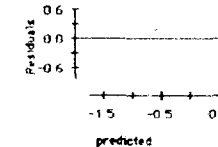
Source	RSS	df	MSS	F
Regression	6.39442	2	3.1972	33.9
Residual	1.60216	17	0.0942	
Total	7.99658	19	0.4209	

 $r^2 = 80.0\%$   $N_{obs} = 20$   $S = 0.3070$ 

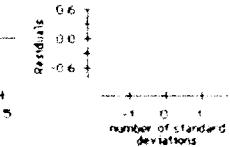
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: BCF PCBs

## Regression Results

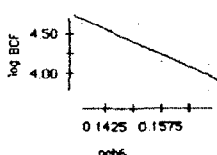
Variable	Coef	Std. Error	t
Constant	7.5756	0.6252	12.1
np6	-21.162	4.009	-5.28

## Analysis of Variance Table

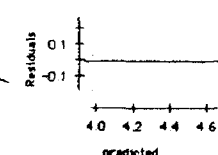
Source	RSS	df	MSS	F
Regression	0.641375	1	0.6414	27.9
Residual	0.138126	6	0.023	
Total	0.779501	7	0.1114	

 $r^2 = 82.3\%$   $N_{obs} = 8$   $S = 0.1517$ 

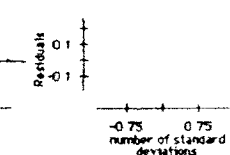
## Log BCF vs. MCI



## Residual vs. Predicted



## Residual vs. Prob.



# STATISTICS Class: Koc Sabljic 1988

Sabljić, A. 1988. Application of Molecular Topology for the Estimation of Physical Data for Environmental Chemicals.

## Regression Results

Variable	Coef	Std. Error	t
constant	43		
np1	53		

## Analysis of Variance Table

Source	RSS	df	MSS	F
Regression				
Residual				
Total				

 $r^2 = .877$   $N_{obs} = 72$   $S = 0.282$

**STATISTICS Class: Koc Sabljic 1988 P<sub>1</sub>'s**

Sabljic, A. 1988 Application of Molecular Topology for the Estimation of Physical Data for Environmental Chemicals.

**Regression Results**

Variable	Coef.	Std. Error	t
constant	0.10		
np1	0.63		

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression				
Residual				

$r^2 = 986$   $n_{obs} = 8$   $S = 0.202$

**STATISTICS Class: S Gerstl 1987 Alcohols**

Gerstl, Z. and C.S. Helling. 1987 Evaluation of molecular connectivity as a predictive method for the adsorption of pesticides by soils. J. Environ. Sci. Health, B22(1), 55-69

**Regression Results**

Variable	Coef.	Std. Error	t
constant	8.37		
vp1	1.18		

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression				
Residual				

$r^2 = 998$   $n_{obs} = 7$   $S =$

**STATISTICS Class: Kow Gerstl 1987 Alcohols**

Gerstl, Z. and C.S. Helling. 1987. Evaluation of molecular connectivity as a predictive method for the adsorption of pesticides by soils. J. Environ. Sci. Health, B22(1), 55-69

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-0.56		
vp1	0.467		

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression				
Residual				

$r^2 =$   $n_{obs} =$   $S =$

**STATISTICS Class: S Alcohols [3 MCIs]**

Nirmalakhandan, N.N. and R.E. Speece. 1988 Prediction of Aqueous Solubility of Organic Chemicals Based on Molecular Structure. Environ. Sci. Technol. 22:3 p328-337

**Regression Results**

Variable	Coef.	Std. Error	t
constant	4.52		
np1	-4.018		
vp1	2.905		
vp3	0.185		

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression				
Residual				

$r^2 = 980$   $n_{obs} = 50$   $S =$

**STATISTICS Class: S Universal from Kow**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.73309	0.1511	4.85
log Kow	-1.1348	0.0388	-29.2
MP-25	-0.01		

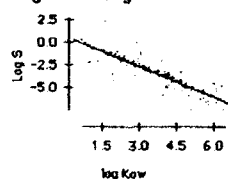
**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression	442.425	1	442	853
Residual	89.6957	173	0.518472	

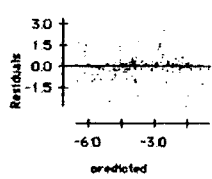
$r^2 = 83.1\%$   $n_{obs} = 175$   $S = 0.7201$

Results in units of mol/L

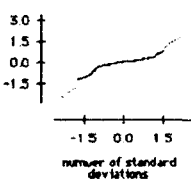
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S Alcohols from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.7476	0.0981	7.63
log Kow	-1.0573	0.0388	-27.2
MP-25	-0.01		

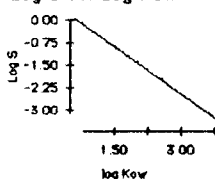
**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression	6.73479	1	6.7348	741
Residual	0.036340	4	0.009085	

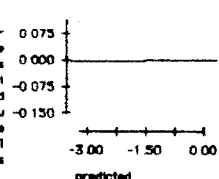
$r^2 = 99.5\%$   $n_{obs} = 6$   $S = 0.0953$

Results in units of mol/L

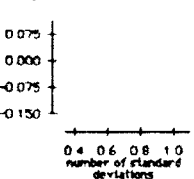
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S Anilines from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.0026	0.2687	0.01
log Kow	-0.0238	0.1045	-8.84
MP-25	-0.01		

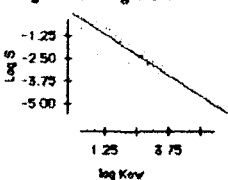
**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression	20.4480	1	20.44	78.1
Residual	4.46113	17	0.261831	

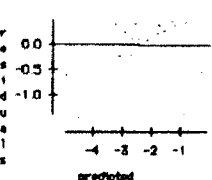
$r^2 = 82.1\%$   $n_{obs} = 19$   $S = 0.5117$

Results in units of mol/L

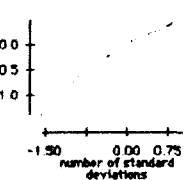
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S Carbamates from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.0187	0.4552	0.041
log Kow	-1.0384	0.1653	-6.69
MP-25	-0.01		

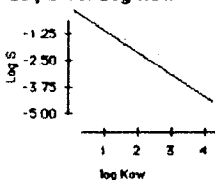
**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression	12.6937	1	12.69	44.7
Residual	1.41978	5	0.283956	

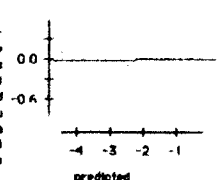
$r^2 = 89.0\%$   $n_{obs} = 7$   $S = 0.5329$

Results in units of mol/L

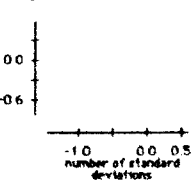
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.



**STATISTICS Class: S Halogenated Aliphatics from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.7356	0.3293	2.23
log Kow	-1.1699	0.0893	-13.1
MP-25	-0.01		

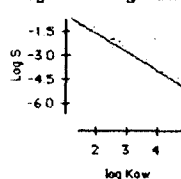
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	79.0886	1	79.0886	171
Residual	16.3297	35	0.466563	

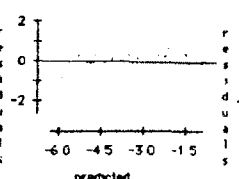
 $r^2 = 83.0\%$   $N_{obs} = 37$   $S = 0.6831$ 

Results in units of mol/L

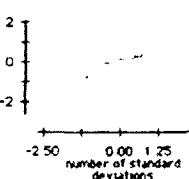
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S Halogenated Aromatics from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.4925	0.289	1.7
log Kow	-1.0829	0.0703	-15.4
MP-25	-0.01		

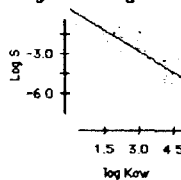
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	134.264	1	134.264	237
Residual	32.2310	57	0.565456	

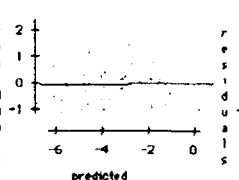
 $r^2 = 80.6\%$   $N_{obs} = 59$   $S = 0.7520$ 

Results in units of mol/L

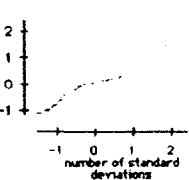
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S PAHs from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	-1.1902	0.5578	-2.13
log Kow	-0.7905	0.1162	-6.81
MP-25	-0.01		

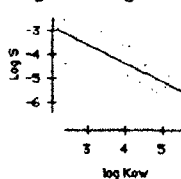
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	14.4474	1	14.4474	46.3
Residual	5.30311	17	0.311948	

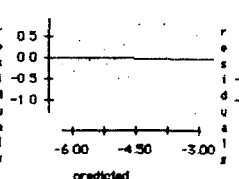
 $r^2 = 73.1\%$   $N_{obs} = 19$   $S = 0.5585$ 

Results in units of mol/L

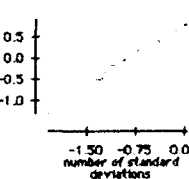
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: Kow Universal from S & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	1.14856	0.0965	12.0
log S (mol/L)	-0.73268	0.0261	-29.2
MP-25	-0.07327		

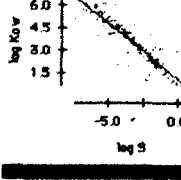
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	285.65	1	285.65	853
Residual	57.9118	173	0.33476	

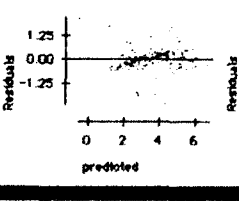
 $r^2 = 83.1\%$   $N_{obs} = 176$   $S = 0.5786$ 

Results in units of

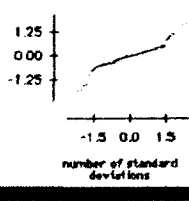
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S Nonhalogenated Aliphatics from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.7356	0.3293	2.23
log Kow	-1.1699	0.0893	-13.1
MP-25	-0.01		

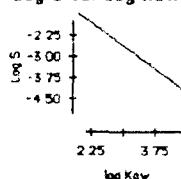
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	11.3747	1	11.3747	5.6
Residual	0.97086	10	0.097086	

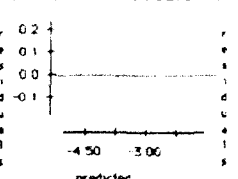
 $r^2 = 59.1\%$   $N_{obs} = 10$   $S = 0.1107$ 

Results in units of mol/L

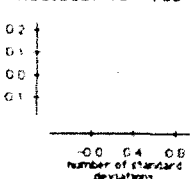
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S PCBs from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	1.9217	0.767	2.51
log Kow	-1.3791	0.1433	-9.62
MP-25	-0.01		

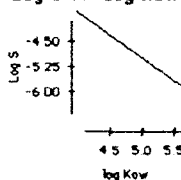
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	5.76095	1	5.76095	92.6
Residual	0.373253	6	0.0622088	

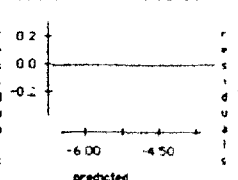
 $r^2 = 91.9\%$   $N_{obs} = 8$   $S = 0.2494$ 

Results in units of mol/L

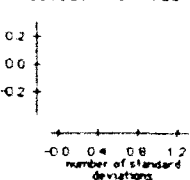
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: S Ureas from Kow & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	-4.126	0.1916	-2.15
log Kow	-0.9512	0.1011	-9.41
MP-25	-0.01		

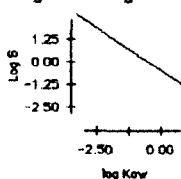
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	17.2055	1	17.2055	88.6
Residual	0.97241	5	0.194482	

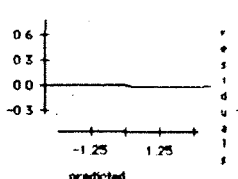
 $r^2 = 94.7\%$   $N_{obs} = 7$   $S = 0.4407$ 

Results in units of mol/L

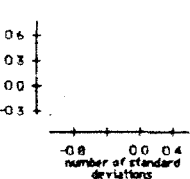
Log S vs. Log Kow



Residual vs. Predicted



Residual vs. Prob.


**STATISTICS Class: Kow Alcohols from S & Mp**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.716	0.0694	10.3
log S (mol/L)	-0.9407	0.0345	-27.2
MP-25	-0.009407		

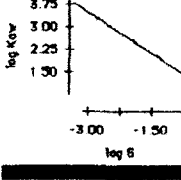
**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression	5.99175	1	5.99175	741
Residual	0.032331	4	0.008083	

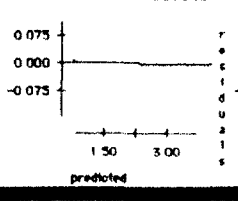
 $r^2 = 99.5\%$   $N_{obs} = 6$   $S = 0.0899$ 

Results in units of

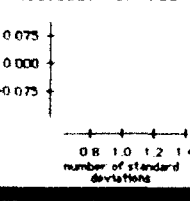
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.





## STATISTICS Class: Kow Anilines from S &amp; Mp

Regression Results

Variable	Coef.	Std. Error	t
Constant	0.4119	0.2418	1.7
log S (mol/L)	-0.8889	0.1005	-8.84
MP-25	-0.008889		

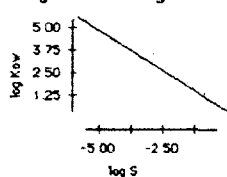
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	19.6729	1	19.67	78.1
Residual	4.28305	17	0.251944	

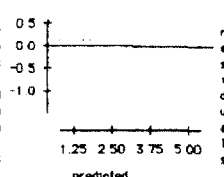
 $r^2 = 82.1\%$   $R_{adj} = 19$   $S = 0.5019$ 

Results in units of

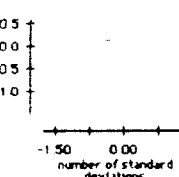
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Kow Halogenated Aliphatics from S &amp; Mp

Regression Results

Variable	Coef.	Std. Error	t
Constant	1.1097	0.2	5.55
log S (mol/L)	-0.7099	0.0542	-13.1
MP-25	-0.007099		

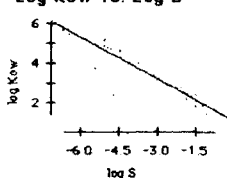
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	48.5377	1	48.54	171
Residual	9.90900	36	0.283114	

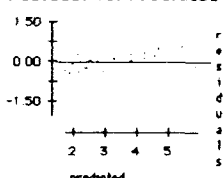
 $r^2 = 83.0\%$   $R_{adj} = 37$   $S = 0.5321$ 

Results in units of

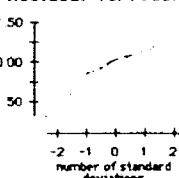
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Kow Halogenated Aromatics from S &amp; Mp

Regression Results

Variable	Coef.	Std. Error	t
Constant	1.1158	0.1963	5.68
log S (mol/L)	-0.7447	0.0483	-15.4
MP-25	-0.007447		

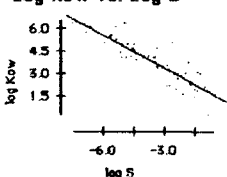
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	92.3245	1	92.32	237
Residual	22.1632	57	0.388828	

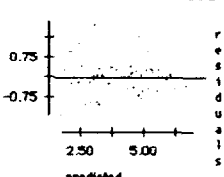
 $r^2 = 80.6\%$   $R_{adj} = 59$   $S = 0.6236$ 

Results in units of

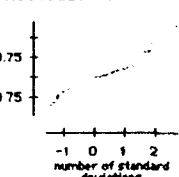
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Koc Halogenated Aromatics from S

Regression Results

Variable	Coef.	Std. Error	t
Constant	0.2786	0.2029	1.37
log S (mol/L)	-0.6063	0.0442	-13.7

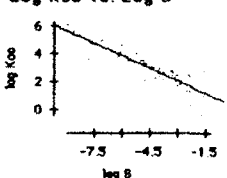
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	61.8776	1	61.88	188
Residual	17.1201	52	0.329406	

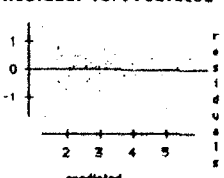
 $r^2 = 78.3\%$   $R_{adj} = 54$   $S = 0.5739$ 

Results in units of L/kg

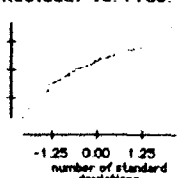
Log Koc vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Kow Carbamates from S &amp; Mp

Regression Results

Variable	Coef.	Std. Error	t
Constant	0.2806	0.3964	0.708
log S (mol/L)	-0.8662	0.1295	-6.69
MP-25	-0.008662		

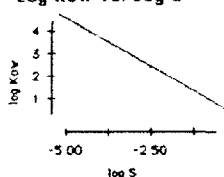
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	10.5891	1	10.59	44.7
Residual	1.18438	7	0.169196	

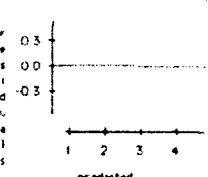
 $r^2 = 89.4\%$   $R_{adj} = 7$   $S = 0.4167$ 

Results in units of

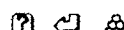
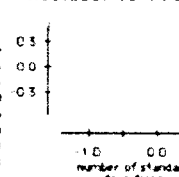
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Kow Nonhalogenated Aliphatics from S &amp; Mp

Regression Results

Variable	Coef.	Std. Error	t
Constant	0.6879	0.1116	6.17
log S (mol/L)	-0.8882	0.0291	-30.5
MP-25	-0.008882		

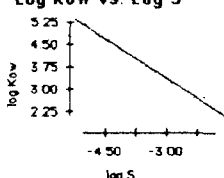
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	9.05040	1	9.0504	679
Residual	0.077964	8	0.009746	

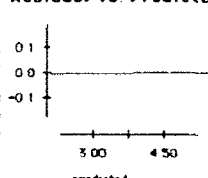
 $r^2 = 99.1\%$   $R_{adj} = 10$   $S = 0.0987$ 

Results in units of

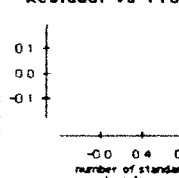
Log Kow vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Koc Universal from S

Regression Results

Variable	Coef.	Std. Error	t
Constant	0.956372	0.1653	5.78
log S (mol/L)	-0.58564	0.0421	-13.9
MP-25	-0.005856		

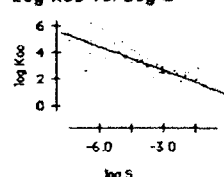
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	70.0563	1	70.06	194
Residual	23.8365	66	0.36119	

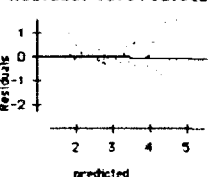
 $r^2 = 74.6\%$   $R_{adj} = 68$   $S = 0.601$ 

Results in units of L/kg

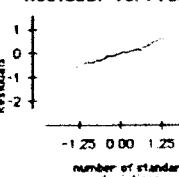
Log Koc vs. Log S



Residual vs. Predicted



Residual vs. Prob.



## STATISTICS Class: Koc PAHs from S

Regression Results

Variable	Coef.	Std. Error	t
Constant	0.6228	0.6433	1.15
log S (mol/L)	-0.6428	0.0914	-7.04

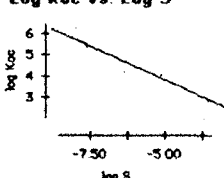
Analysis of Variance Table

Source	RSS	df	MSS	F
Regression	14.4440	1	14.44	49.5
Residual	2.62593	9	0.291770	

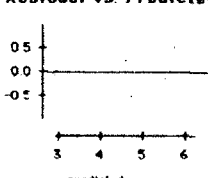
 $r^2 = 84.6\%$   $R_{adj} = 11$   $S = 0.5402$ 

Results in units of L/kg

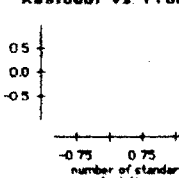
Log Koc vs. Log S



Residual vs. Predicted



Residual vs. Prob.

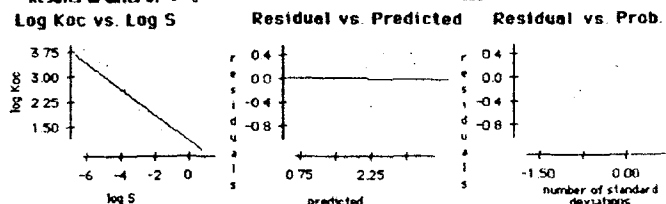




**STATISTICS Class: Koc Ureas from S**

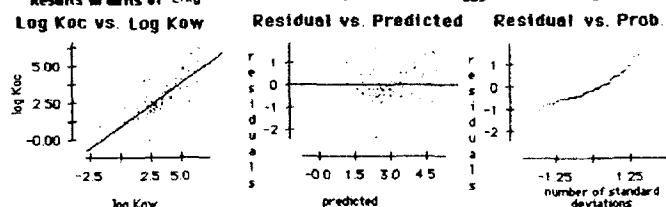
Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	1.1216	0.239	4.69	Regression	6.93911	1	6.93911	37.5
log S (mol/L)	-.3011	0.0639	-6.13	Residual	2.58898	14	0.184927	

Results in units of L/kg  $r^2 = 72.8\%$   $N_{obs} = 16$   $S = 0.4300$


**STATISTICS Class: Koc Universal from Kow**

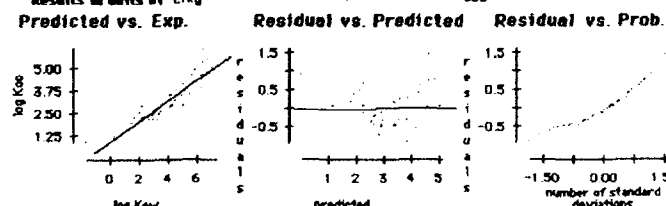
Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	1.0001	0.1723	5.81	Regression	93.7441	1	93.7441	175
log Kow	0.5881	0.0445	13.2	Residual	42.8715	80	0.535803	

Results in units of L/kg  $r^2 = 68.6\%$   $N_{obs} = 82$   $S = 0.7320$


**STATISTICS Class: Koc Halogenated Aromatics from Kow**

Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	0.9838	0.19	5.18	Regression	42.9850	1	42.9850	161
log Kow	0.5723	0.0451	12.7	Residual	9.85704	37	0.266407	

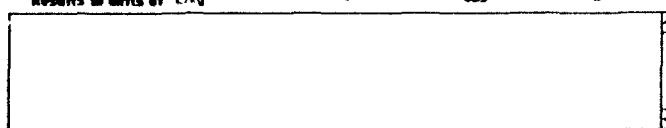
Results in units of L/kg  $r^2 = 81.3\%$   $N_{obs} = 39$   $S = 0.5161$


**STATISTICS Class: Koc Karickhoff et al. 1979 from Kow**

Karickhoff, S.W., D.S. Brown, and T.A. Scott, 1979. Sorption of hydrophilic pollutants on natural sediments. Water Res. 13:241-248.

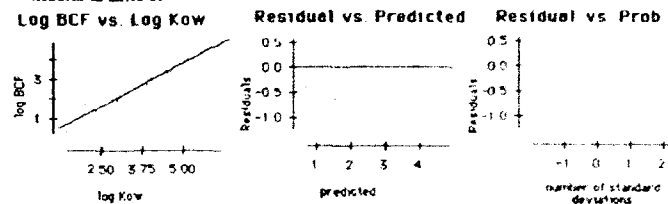
Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
constant	-.21							
log Kow	1.0							

Results in units of L/kg  $r^2 =$   $N_{obs} =$   $S =$


**STATISTICS Class: BCF Universal from Kow**

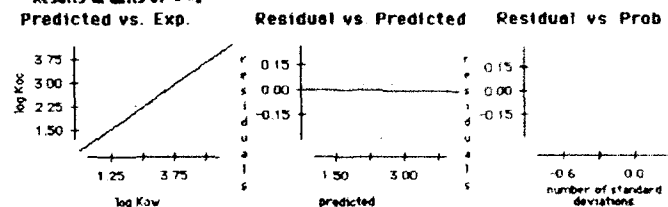
Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	-0.58027	0.1978	-2.93	Regression	46.6153	1	46.6153	3.11
log Kow	0.880365	0.0484	18.2	Residual	6.51802	46	0.14170	

Results in units of  $r^2 = 87.8\%$   $N_{obs} = 48$   $S = 0.364$


**STATISTICS Class: Koc Anilines from Kow**

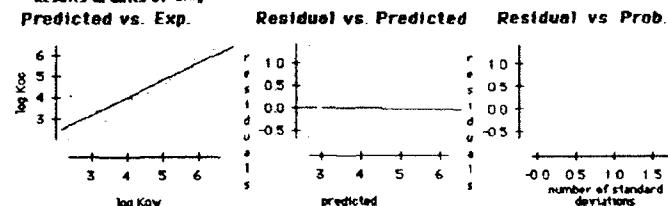
Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	0.8605	0.177	4.86	Regression	4.30510	1	4.30510	91.8
log Kow	0.5655	0.059	9.58	Residual	0.335126	7	0.047875	

Results in units of L/kg  $r^2 = 92.9\%$   $N_{obs} = 9$   $S = 0.2168$


**STATISTICS Class: Koc PAHs from Kow**

Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	0.7968	0.638	1.14	Regression	14.1748	1	14.1748	38.4
log Kow	0.6231	0.1328	6.2	Residual	2.94042	8	0.367552	

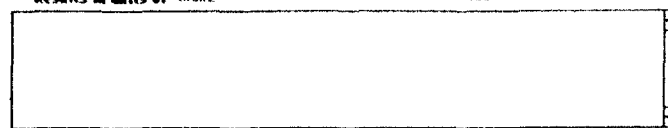
Results in units of L/kg  $r^2 = 92.8\%$   $N_{obs} = 10$   $S = 0.6043$


**STATISTICS Class: S Hansch et al. 1968 from Kow**

Hansch, C., J.E. Quinn, and G.L. Lawrence, "The Linear Free-Energy Relationships between Partition Coefficients and the Aqueous Solubility of Organic Liquids," J. Org. Chem., 44, 347-50 (1968)

Regression Results				Analysis of Variance Table				
Variable	Coef.	Std. Error	t	Source	RSS	df	MSS	F
Constant	0.978							
Log Kow	-1.339							

Results in units of mol/L  $r^2 = 87.4$   $N_{obs} = 156$   $S =$



**STATISTICS Class:** S Saeger et al. 1979 Phosphate esters from

Saeger, V.W., O. Hicks, R.G. Kaley, P.R. Michael, J.P. Mieux and E.S. Tucker, "Environmental Fate of Selected Phosphate Esters," Environ. Sci. Technol., 13, 840-44 (1979).

**Regression Results**

Variable	Coef.	Std. Error	t
Constant	12.9		
Log Kow	-2.38		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of  $\mu\text{mol/L}$ 
 $r^2 = .656$   $N_{\text{obs}} = 11$   $S =$ 
**STATISTICS Class:** S Yalkowsky et al. 1979 Halobenzenes from

Yalkowsky, S.H., R.J. Orr and S.C. Valvani, "Solubility and Partitioning: 3. The Solubility of Halobenzenes in Water," Ind. Eng. Chem. Fundam. 18, 351-63 (1979).

**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.7178		
Log Kow	0.0874		
Mp	0.0095		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of  $\text{mol/L}$ 
 $r^2 = .990$   $N_{\text{obs}} = 15$   $S =$ 
**STATISTICS Class:** S Mackay and Shiu 1977 PAHs from Kow&Mp

Mackay, D. and W.Y. Shiu, "Aqueous Solubility of Polynuclear Aromatic Hydrocarbons," J. Chem. Eng. Data, 11, 399-402 (1979).

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-0.012		
Log Kow	-0.88		
Mp	-0.01		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of  $\text{mol/L}$ 
 $r^2 = .979$   $N_{\text{obs}} = 32$   $S =$ 
**STATISTICS Class:** Koc Kenaga and Goring 1978 Pesticides

Kenaga, E.E. and C.A.I. Goring, "Relationship Between Water Solubility, Soil-Sorption, Octanol-Water Partitioning and Bioconcentration of Chemicals in Biota," pre-publication copy of paper dated October 13.

**Regression Results**

Variable	Coef.	Std. Error	t
constant	3.64		
log S $\text{mg/L}$	-0.55		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of

 $r^2 = .714$   $N_{\text{obs}} = 106$   $S =$ 
**STATISTICS Class:** Koc Kenaga & Goring 1978 Pesticides from

Kenaga, E.E. and C.A.I. Goring, "Relationship Between Water Solubility, Soil-Sorption, Octanol-Water Partitioning and Bioconcentration of Chemicals in Biota," pre-publication copy of paper dated October 13.

**Regression Results**

Variable	Coef.	Std. Error	t
Constant	1.377		
log Kow	.544		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of

 $r^2 = .74$   $N_{\text{obs}} = 45$   $S =$ 
**STATISTICS Class:** S Kenaga and Goring from Kow

Kenaga, E.E. and C.A.I. Goring, "Relationship Between Water Solubility, Soil-Sorption, Octanol-Water Partitioning and Bioconcentration of Chemicals in Biota," pre-publication copy of paper dated October 13.

**Regression Results**

Variable	Coef.	Std. Error	t
constant	4.184		
Log Kow	-.922		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of  $\text{mg/L}$ 
 $r^2 = .740$   $N_{\text{obs}} = 90$   $S =$ 
**STATISTICS Class:** S Yalkowsky PAHs from Kow

Yalkowsky, S.H. and S.C. Valvani, "Solubilities and partitioning: 2. Relationships between aqueous solubilities, partition coefficients and molecular surface areas of rigid aromatic hydrocarbons," J. Chem. Eng. Data,

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-0.012		
Log Kow	-0.88		
Mp	-.01		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of  $\text{mol/L}$ 
 $r^2 =$   $N_{\text{obs}} =$   $S =$ 
**STATISTICS Class:** S Chiou Halo. Hydrocarbons

Chiou, C.T., and V.H. Freed (technical note), Environ. Sci. Technol., 11, 1219-20 (1977)

**Regression Results**

Variable	Coef.	Std. Error	t
constant	6.5		
Log Kow	-.962		

**Analysis of Variance Table**

Source	RSS	df	MSS	F

Results in units of  $\mu\text{mol/L}$ 
 $r^2 = .678$   $N_{\text{obs}} = 9$   $S =$ 

Halogenated 1- and 2-carbon hydrocarbons (8 with Cl, 1 with Br)

**STATISTICS Class: Koc Brown et al aromatics**

Brown, D.S., S.W. Karickhoff and E.W. Flagg, "Empirical Prediction of Organic Pollutant Sorption in Natural Sediments," to be submitted to J. Environ. Qual.

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-0.006		
Log Kow	0.937		

**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression				
Residual				

Results in units of

 $r^2 = .95$   $n_{obs} = 14$   $S =$ 

Aromatics, polynuclear aromatics, triazines and dinitroaniline herbicides

**STATISTICS Class: Koc Briggs 1973**

Briggs, G. G., "A Simple Relationship Between Soil Adsorption of Organic Chemicals and Their Octanol/Water partition coefficients," Proc 7th British Insecticide and Fungicide Conf. Vol 1, The Boots Company Ltd.

**Regression Results**

Variable	Coef.	Std. Error	t
constant	.856		
Log Kow	5.24		

**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression				
Residual				

Results in units of

 $r^2 = .84$   $n_{obs} = 30$   $S =$ 

Substituted phenylureas and alkyl-N-phenylcarbamates

**STATISTICS Class: BCF Kenaga and Goring from S**

Kenaga, E.E. and C.A.I. Goring, "Relationship Between Water Solubility, Soil-Sorption, Octanol-Water Partitioning and Bioconcentration of Chemicals in Biota," pre-publication copy of paper dated October 13,

**Regression Results**

Variable	Coef.	Std. Error	t
constant	2.791		
Log S	-0.564		

**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression				
Residual				

Results in units of

 $r^2 = .49$   $n_{obs} = 35$   $S =$ 

Wide variety of compounds

**STATISTICS Class: S Universal**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	0.9805	0.2381	4.03
TSA	-0.0234	0.0012	-19.6
N-TSA	0.0692	0.0174	3.98
O-TSA	0.0548	0.0057	9.56
Am-TSA	0.0838	0.0152	5.52

**Analysis of Variance Table**

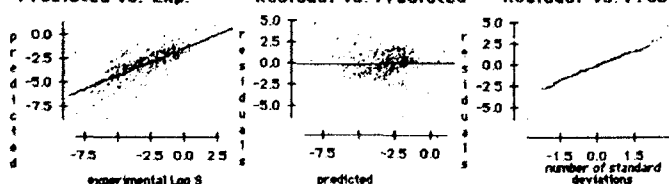
Source	RSS	d <sub>f</sub>	MSS	F
Regression	628.031	4	157	108
Residual	455.707	307	1.48430	

 $r^2 = 58.0\%$   $n_{obs} = 312$   $S = 1.218$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.


**STATISTICS Class: Koc Brown 1979 Triazines**

Brown, D.S. U.S. Environmental Protection Agency, Athens, GA, personal communication September 20, 1979. Published in "Handbook of Chemical Property Estimation Methods" W.J. Lyman, W.F. Reehl, and D.H. Rosenblatt

**Regression Results**

Variable	Coef.	Std. Error	t
constant	0.02		
Log Kow	0.4		

**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression				
Residual				

Results in units of

 $r^2 =$   $n_{obs} = 4$   $S =$ 

s-Triazines and dinitroaniline herbicides

**STATISTICS Class: BCF Veith et al from Kow**

Veith, G.D., K.J. Macek, S. R. Petocelli and J. Caroll, "An Evaluation of Using Partition Coefficients and Water Solubility to Estimate Bioconcentration Factors for Organic Chemicals in Fish," J. Fish. Res. Board Can. (1980)

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-0.23		
Log Kow	0.76		

**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression				
Residual				

Results in units of

 $r^2 = .873$   $n_{obs} = 14$   $S =$ 

Wide variety of compounds

**STATISTICS Class: H PWS\*RT**
**Regression Results**

Variable	Coef.	Std. Error	t
constant	-6.394		
log P <sub>v</sub> (Pa)	1.0		
log S (mol/L)	-1.0		

**Analysis of Variance Table**

Source	RSS	d <sub>f</sub>	MSS	F
Regression				
Residual				

Results in units of dimensionless

 $r^2 =$   $n_{obs} =$   $S =$ 
**STATISTICS Class: S Alcohols**
**Regression Results**

Variable	Coef.	Std. Error	t
Constant	4.1567	0.1817	22.9
TSA	-0.0276	0.0005	-54.4
O-TSA	-0.0578	0.0152	-3.8
MP-25	-0.01		

**Analysis of Variance Table**

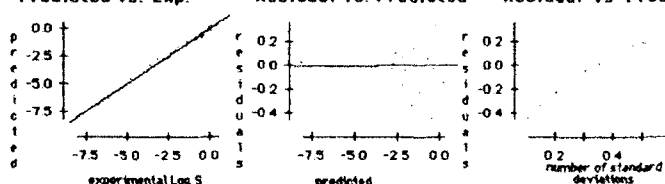
Source	RSS	d <sub>f</sub>	MSS	F
Regression	147.398	2	73.7	1521
Residual	1.40529	29	0.04845	

 $r^2 = 99.1\%$   $n_{obs} = 32$   $S = 0.2201$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: S Halogenated Aliphatics

## Regression Results

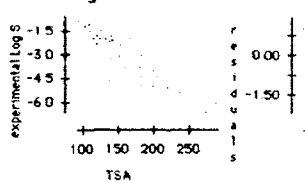
Variable	Coef.	Std. Error	t
Constant	1.7144	0.3573	4.8
TSA	-0.0306	0.0021	-14.5
MP-25	-0.01		

## Analysis of Variance Table

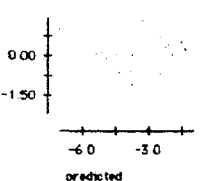
Source	RSS	d <sub>f</sub>	MSS	F
Regression	82.6090	1	82.61	209
Residual	13.7893	35	0.39398	

 $r^2 = 85.7\%$   $n_{obs} = 37$   $S = 0.6277$ 

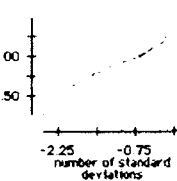
## Log S vs. TSA



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: S PCBs

## Regression Results

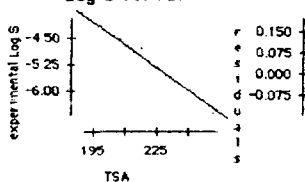
Variable	Coef.	Std. Error	t
Constant	4.3776	0.527	8.31
TSA	-0.0432	0.0023	-18.8
MP-25	-0.01		

## Analysis of Variance Table

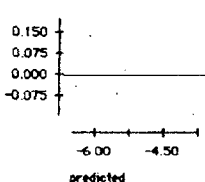
Source	RSS	d <sub>f</sub>	MSS	F
Regression	6.03013	1	6.0301	348
Residual	0.104071	6	0.01734	

 $r^2 = 98.3\%$   $n_{obs} = 8$   $S = 0.1317$ 

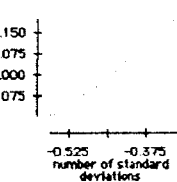
## Log S vs. TSA



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: S Phenols

## Regression Results

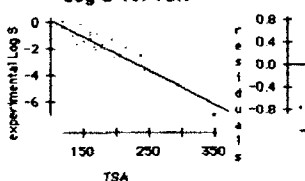
Variable	Coef.	Std. Error	t
Constant	2.9177	0.498	5.86
TSA	-0.0261	0.0027	-9.68
MP-25	-0.01		

## Analysis of Variance Table

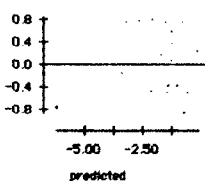
Source	RSS	d <sub>f</sub>	MSS	F
Regression	35.8815	1	35.88	93.7
Residual	6.85000	18	0.38055	

 $r^2 = 93.9\%$   $n_{obs} = 20$   $S = 0.6169$ 

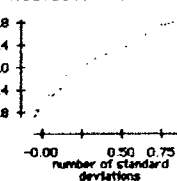
## Log S vs. TSA



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: Kow Universal

## Regression Results

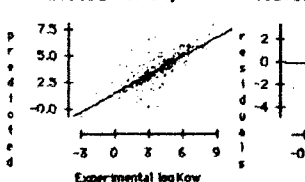
Variable	Coef.	Std. Error	t
Constant	0.3022	0.2431	1.24
TSA	0.0194	0.0012	16.8
N-TSA	-0.1223	0.025	-4.9
O-TSA	-0.052	0.0082	-6.35
AN-TSA	-0.05	0.02	-2.48

## Analysis of Variance Table

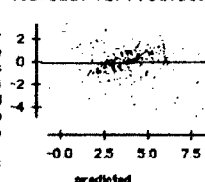
Source	RSS	d <sub>f</sub>	MSS	F
Regression	608.993	4	127	94.7
Residual	352.013	283	1.33845	

 $r^2 = 69.0\%$   $n_{obs} = 288$   $S = 1.167$ 

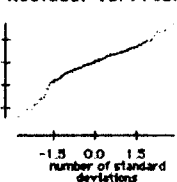
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: S Nonhalogenated Aliphatics

## Regression Results

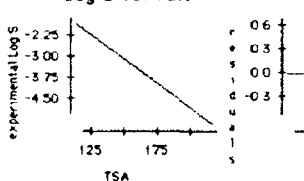
Variable	Coef.	Std. Error	t
Constant	2.047	0.7338	2.79
TSA	-0.0343	0.0043	-7.92
MP-25	0.01		

## Analysis of Variance Table

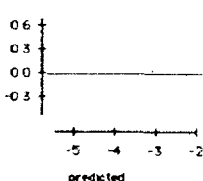
Source	RSS	d <sub>f</sub>	MSS	F
Regression	10.1745	1	10.17	62.7
Residual	1.29817	8	0.16222	

 $r^2 = 88.7\%$   $n_{obs} = 10$   $S = 0.4028$ 

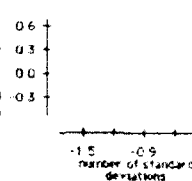
## Log S vs. TSA



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: S PAHs

## Regression Results

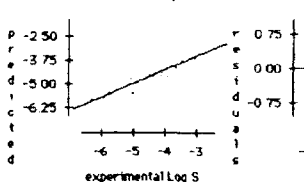
Variable	Coef.	Std. Error	t
Constant	-3.199	0.5198	-6.16
TSA	-0.0216	0.0024	-8.94
O-TSA	0.17	0.022	7.73
MP-25	-0.01		

## Analysis of Variance Table

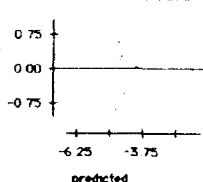
Source	RSS	d <sub>f</sub>	MSS	F
Regression	31.5758	2	15.79	53.9
Residual	11.1357	38	0.29304	

 $r^2 = 73.9\%$   $n_{obs} = 41$   $S = 0.5413$ 

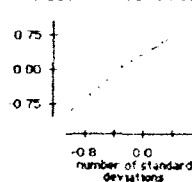
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: S Ureas

## Regression Results

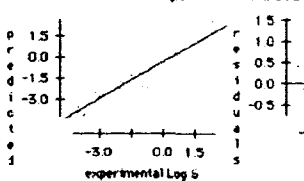
Variable	Coef.	Std. Error	t
Constant	5.8943	0.8787	6.71
TSA	-0.0293	0.0033	-8.52
O-TSA	-0.075	0.0258	-2.93
MP-25	-0.01		

## Analysis of Variance Table

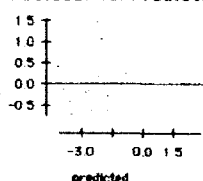
Source	RSS	d <sub>f</sub>	MSS	F
Regression	37.0570	2	18.53	45.2
Residual	5.33016	13	0.41001	

 $r^2 = 87.4\%$   $n_{obs} = 16$   $S = 0.6403$ 

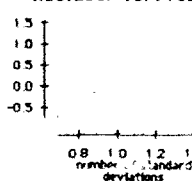
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: Kow Alcohols

## Regression Results

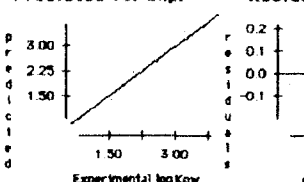
Variable	Coef.	Std. Error	t
Constant	-2.7111	0.4591	-5.9
TSA	0.0256	0.0024	10.5
O-TSA	0.0389	0.0154	2.53

## Analysis of Variance Table

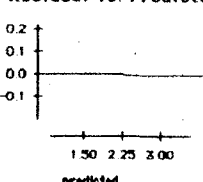
Source	RSS	d <sub>f</sub>	MSS	F
Regression	5.88264	2	2.9413	82.3
Residual	0.141541	3	0.04718	

 $r^2 = 97.7\%$   $n_{obs} = 6$   $S = 0.2172$ 

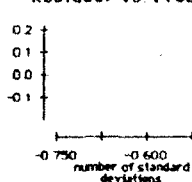
## Predicted vs. Exp.



## Residual vs. Predicted



## Residual vs. Prob



# STATISTICS Class: Kow Carbamates

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-3.0312	1.484	-2.04
TSA	0.0215	0.0055	3.89

## Analysis of Variance Table

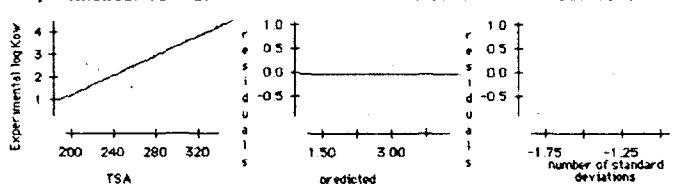
Source	RSS	df	MSS	F
Regression	8.84905	1	8.84905	15.1
Residual	2.92444	5	0.58489	

 $r^2 = 75.2\%$   $n_{obs} = 7$   $S = 0.7648$ 

Experimental vs. TSA

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Nonhalogenated Aliphatics

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.1508	0.8559	-1.75
TSA	0.0205	0.0039	7.9

## Analysis of Variance Table

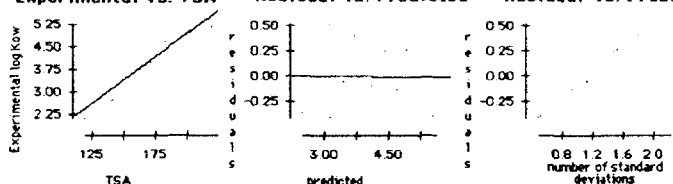
Source	RSS	df	MSS	F
Regression	8.09101	1	8.09101	62.4
Residual	1.03735	8	0.12966	

 $r^2 = 88.6\%$   $n_{obs} = 10$   $S = 0.3801$ 

Experimental vs. TSA

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow PCBs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-2.628	0.3757	-6.99
TSA	0.025	0.0016	16.7

## Analysis of Variance Table

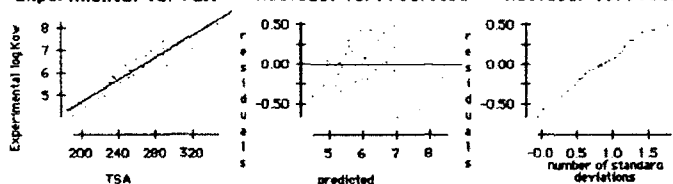
Source	RSS	df	MSS	F
Regression	23.4248	1	23.42	278
Residual	2.52769	30	0.08425	

 $r^2 = 90.3\%$   $n_{obs} = 32$   $S = 0.2903$ 

Experimental vs. TSA

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Phenols

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-0.0307	0.6801	-0.056
TSA	0.017	0.0031	5.43

## Analysis of Variance Table

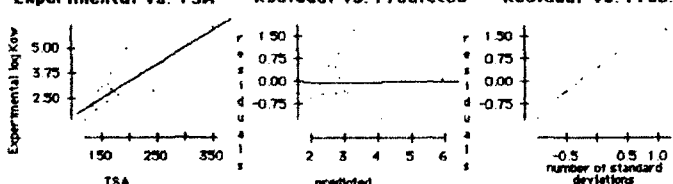
Source	RSS	df	MSS	F
Regression	13.6070	1	13.61	29.5
Residual	7.39176	16	0.46198	

 $r^2 = 64.8\%$   $n_{obs} = 18$   $S = 0.6797$ 

Experimental vs. TSA

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Halogenated Aliphatics

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-2.956	0.2953	-10.1
TSA	0.0226	0.0017	13.1

## Analysis of Variance Table

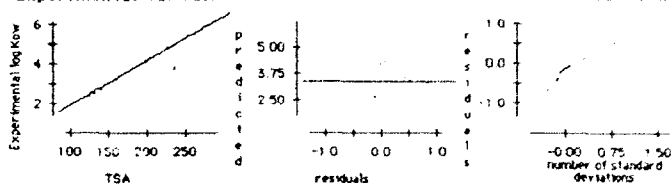
Source	RSS	df	MSS	F
Regression	4.79109	1	4.791	17.3
Residual	9.98576	38	0.27338	

 $r^2 = 82.8\%$   $n_{obs} = 39$   $S = 0.5267$ 

Experimental vs. TSA

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Halogenated Aromatics

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.6555	0.3295	1.99
TSA	0.0208	0.0014	14.3
N-TSA	-1.497	0.0276	-5.43
O-TSA	-0.584	0.0095	-6.16
ArN-TSA	-0.559	0.0177	-3.73

## Analysis of Variance Table

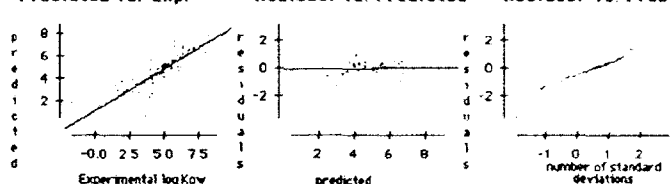
Source	RSS	df	MSS	F
Regression	248.812	4	62.2	79.6
Residual	81.3114	104	0.78184	

 $r^2 = 75.4\%$   $n_{obs} = 108$   $S = 0.8842$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow PAHs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.821	0.7424	-2.18
TSA	0.0243	0.0034	7.1
O-TSA	-1.243	0.0255	-4.88

## Analysis of Variance Table

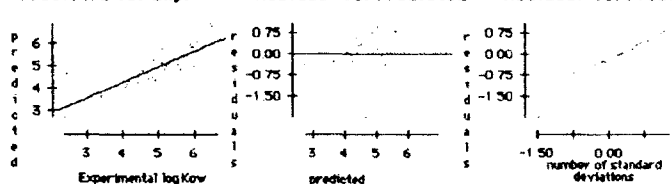
Source	RSS	df	MSS	F
Regression	22.0177	2	11.01	26.7
Residual	9.47765	23	0.41207	

 $r^2 = 69.9\%$   $n_{obs} = 25$   $S = 0.6419$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Triazines

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-1.9958	1.3	-1.64
TSA	0.0204	0.0051	3.98
N-TSA	-0.489	0.0116	-4.2

## Analysis of Variance Table

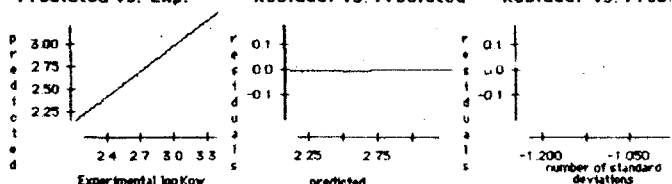
Source	RSS	df	MSS	F
Regression	0.901281	2	0.45064	13.7
Residual	0.085639	2	0.03281	

 $r^2 = 93.2\%$   $n_{obs} = 5$   $S = 0.1812$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



# STATISTICS Class: Kow Ureas

## Regression Results

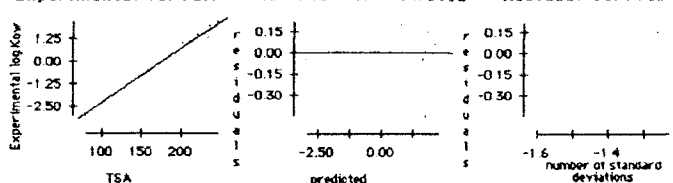
Variable	Coef.	Std. Error	t
Constant	-5.3722	0.3591	-15
TSA	0.0306	0.0017	18.2

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	18.7323	1	18.73	330
Residual	0.284226	5	0.05684	

 $r^2 = 98.5\%$   $n_{obs} = 7$   $S = 0.2384$ 

Experimental vs. TSA      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: Pv Halogenated Aliphatics

## Regression Results

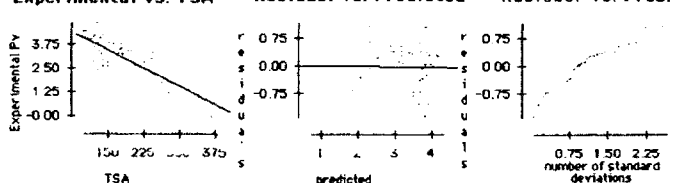
Variable	Coef.	Std. Error	t
Constant	5.4345	0.2932	18.5
TSA	-0.0128	0.0017	-7.74

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	20.1520	1	20.15	59.9
Residual	12.1187	36	0.33663	

 $r^2 = 62.4\%$   $n_{obs} = 38$   $S = 0.5802$ 

Experimental vs. TSA      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: Pv Halogenated Aromatics

## Regression Results

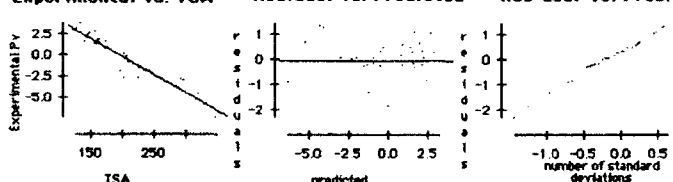
Variable	Coef.	Std. Error	t
Constant	8.3422	0.54	15.4
TSA	-0.0426	0.0026	-16.5

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	182.189	1	182	272
Residual	20.0981	30	0.66993	

 $r^2 = 90.1\%$   $n_{obs} = 32$   $S = 0.8185$ 

Experimental vs. TSA      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: Pv PAHs

## Regression Results

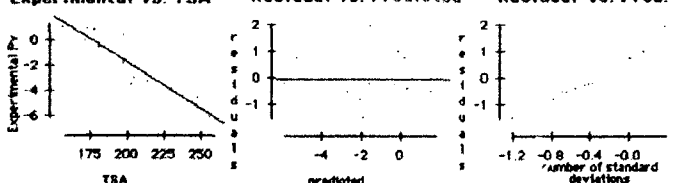
Variable	Coef.	Std. Error	t
Constant	13.1693	1.975	6.66
TSA	-0.0748	0.0099	-7.56

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	54.5428	1	54.54	67.1
Residual	9.54705	10	0.95470	

 $r^2 = 85.1\%$   $n_{obs} = 12$   $S = 0.9771$ 

Experimental vs. TSA      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: Pv Universal

## Regression Results

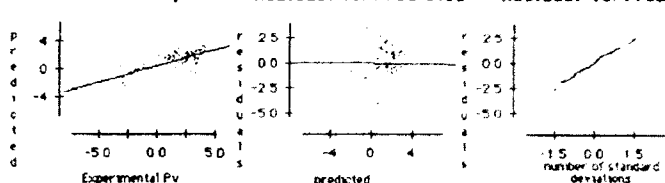
Variable	Coef.	Std. Error	t
Constant	6.506	0.5308	12.3
TSA	-0.0288	0.0029	-10.1
O-TSA	0.08	0.0227	3.52

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	351.366	2	176	56.7
Residual	380.131	126	3.09628	

 $r^2 = 47.4\%$   $n_{obs} = 129$   $S = 1.760$ 

Predicted vs. Exp.      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: Pv Nonhalogenated Aliphatics

## Regression Results

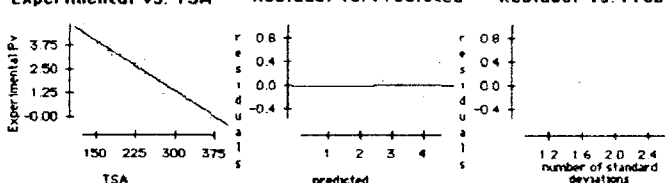
Variable	Coef.	Std. Error	t
Constant	6.7505	0.2911	23.2
TSA	-0.0181	0.0014	-12.9

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	21.2842	1	21.28	188
Residual	2.41075	19	0.12688	

 $r^2 = 89.8\%$   $n_{obs} = 21$   $S = 0.3562$ 

Experimental vs. TSA      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: Pv PCBs

## Regression Results

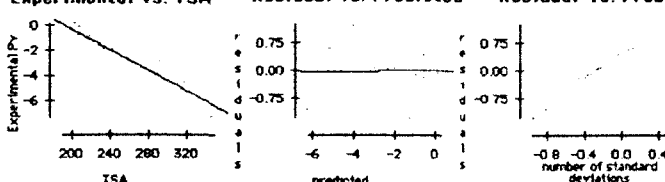
Variable	Coef.	Std. Error	t
Constant	7.9829	0.9298	8.59
TSA	-0.0413	0.0038	-10.9

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	47.2893	1	47.29	120
Residual	5.14347	13	0.39565	

 $r^2 = 90.2\%$   $n_{obs} = 15$   $S = 0.6290$ 

Experimental vs. TSA      Residual vs. Predicted      Residual vs. Prob.



# STATISTICS Class: H Universal

## Regression Results

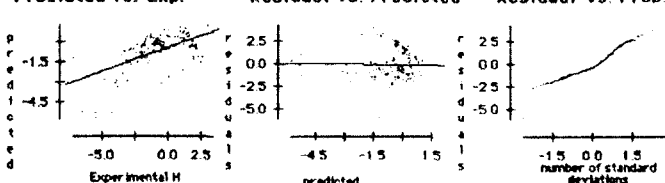
Variable	Coef.	Std. Error	t
Constant	1.8837	0.3749	4.97
TSA	-0.012	0.0018	-6.36
O-TSA	-0.0807	0.0183	-4.97

## Analysis of Variance Table

Source	RSS	d <sub>f</sub>	MSS	F
Regression	207.041	2	104	40.4
Residual	361.505	141	2.56388	

 $r^2 = 36.4\%$   $n_{obs} = 144$   $S = 1.601$ 

Predicted vs. Exp.      Residual vs. Predicted      Residual vs. Prob.



## STATISTICS Class: H Halogenated Aromatics

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.629	0.3242	1.94
TSA	-0.103	0.0015	-6.98
O-TSA	-0.0724	0.0117	-6.19

## Analysis of Variance Table

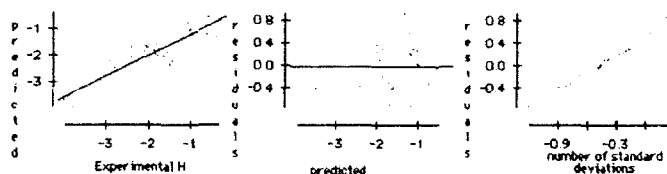
Source	RSS	df	MSS	F
Regression	24.1681	2	12.08	64.7
Residual	5.97456	32	0.18670	

 $r^2 = 80.2\%$   $n_{obs} = 35$   $S = 0.4321$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Koc Halogenated Aromatics

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.1857	0.3588	0.462
TSA	0.0168	0.0016	10.3
N-TSA	-0.0631	0.019	-3.32
O-TSA	-0.052	0.0082	-6.32
A/N-TSA	-0.0518	0.013	-3.97

## Analysis of Variance Table

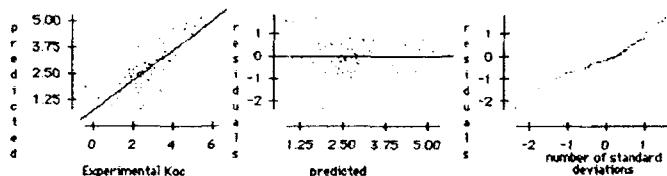
Source	RSS	df	MSS	F
Regression	73.5330	4	18.38	39.5
Residual	29.8175	64	0.46589	

 $r^2 = 71.1\%$   $n_{obs} = 69$   $S = 0.6826$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Koc Triazines

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-0.6948	0.5928	-1.17
TSA	0.0118	0.0021	5.5

## Analysis of Variance Table

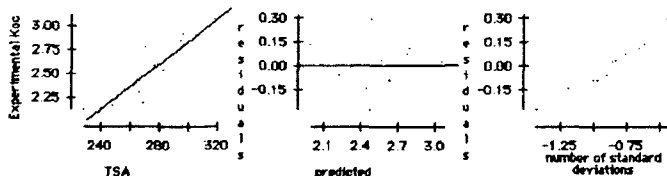
Source	RSS	df	MSS	F
Regression	0.809500	1	0.80950	30.2
Residual	0.240918	9	0.02676	

 $r^2 = 77.1\%$   $n_{obs} = 11$   $S = 0.1636$ 

Experimental vs. TSA

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: BCF Universal

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.0344	0.7002	0.049
TSA	0.0138	0.0031	4.48
N-TSA	-0.292	0.1224	-2.39

## Analysis of Variance Table

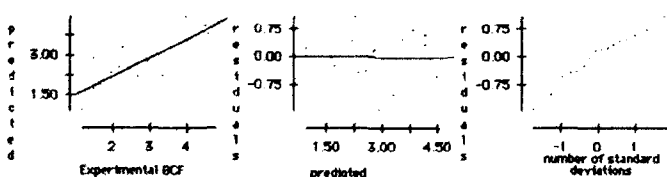
Source	RSS	df	MSS	F
Regression	15.1450	2	7.572	15.5
Residual	6.33814	13	0.48756	

 $r^2 = 70.5\%$   $n_{obs} = 16$   $S = 0.6982$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Koc Universal

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.2239	0.2634	0.884
TSA	0.0145	0.0012	12
N-TSA	-0.0409	0.0134	-3.04
O-TSA	-0.247	0.0049	-5.04
A/N-TSA	0.445	0.0112	3.99

## Analysis of Variance Table

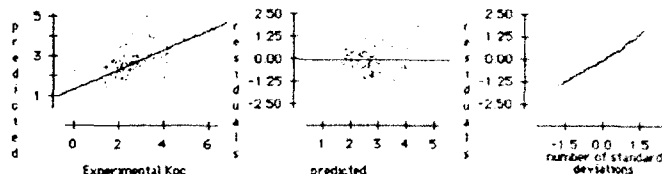
Source	RSS	df	MSS	F
Regression	123.539	4	30.9	39.3
Residual	127.420	162	0.78654	

 $r^2 = 49.2\%$   $n_{obs} = 167$   $S = 0.8869$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Koc PAHs

## Regression Results

Variable	Coef.	Std. Error	t
Constant	0.8402	0.6065	1.39
TSA	-0.1602	0.0182	-8.81
O-TSA	0.0178	0.0028	6.93

## Analysis of Variance Table

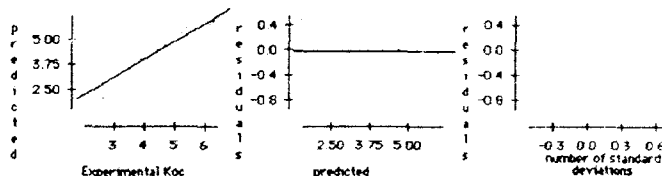
Source	RSS	df	MSS	F
Regression	24.5885	2	12.29	57.4
Residual	2.57223	12	0.21435	

 $r^2 = 90.5\%$   $n_{obs} = 15$   $S = 0.4630$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: Koc Ureas

## Regression Results

Variable	Coef.	Std. Error	t
Constant	-3.1631	1.147	-2.78
TSA	0.0154	0.0028	5.54
N-TSA	0.148	0.0681	2.17
O-TSA	0.0404	0.0149	2.73

## Analysis of Variance Table

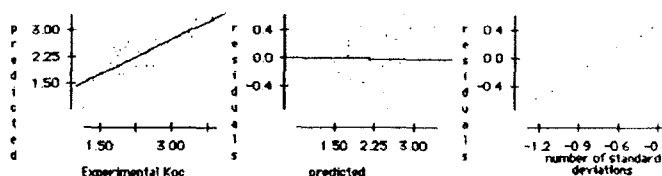
Source	RSS	df	MSS	F
Regression	8.38117	3	2.7937	15.3
Residual	4.02469	22	0.18294	

 $r^2 = 67.6\%$   $n_{obs} = 26$   $S = 0.4277$ 

Predicted vs. Exp.

Residual vs. Predicted

Residual vs. Prob.



## STATISTICS Class: s PAHs from Lit

Fugate, Harry N. 1989. Using Total Molecular Surface Area in Quantitative Structure Activity Relationships to Estimate Environmental Fate and Transport Parameters. Thesis, Utah State University.

## Regression Results

Variable	Coef.	Std. Error	t
constant	3.78		
TSA	-0.05		

## Analysis of Variance Table

Source	RSS	df	MSS	F
Regression				
Residual				

Results in units of moles/L

 $r^2 = .90$   $n_{obs} = 14$   $S = .22$

**STATISTICS Class:s Alcohols from Lit**

Fugate, Harry N. 1989. Using Total Molecular Surface Area in Quantitative Structure Activity Relationships to Estimate Environmental Fate and Transport Parameters. Thesis, Utah State University.

**Regression Results**

Variable	Coef.	Std. Error	t
constant	3.49	.00108	23.91
TSA	-.03		

Results in units of moles/L

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression	28.41	1	28.41	571.73
Residual	1.84	37	.05	
Total	30.25	38		

 $r^2 = .94$      $n_{\text{obs}} = 39$      $S =$ 
**STATISTICS Class:Kow hydrophobic aromatics**

Doucette, W.J., A.W. Andren. 1987. "Correlation of Octanol/Water Partition Coefficients and Total Molecular Surface Area for Highly Hydrophobic Aromatic Compounds." Environmental Science and Technology 21(8):821-824

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-0.142		
TSA	0.0238		

Results in units of

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression				
Residual				
Total				

 $r^2 = 0.624$      $n_{\text{obs}} = 37$      $S =$ 
**STATISTICS Class:Kow Hawker 1988 PCBs**

Hawker, D.W., and D.W. Connell. 1988. Octanol-Water Partition Coefficients of Polychlorinated Biphenyl Congeners. Environ. Sci. Technol. 22:4 p. 382-387

**Regression Results**

Variable	Coef.	Std. Error	t
constant	-2.20		
TSA	0.0341		

Results in units of

**Analysis of Variance Table**

Source	RSS	df	MSS	F
Regression				
Residual				
Total				

 $r^2 = .959$      $n_{\text{obs}} = 46$      $S =$ 

AL  
HAWKER  
TSA  
S  
S